

연결주의와 구성적 구조*

최 훈

머 리 말

천구백팔십년대 초반까지 인지 현상과 마음의 개념에 대한 이론은 현대적인 디지털 컴퓨터 모형에 근거한 규칙 지배적 기호 조작(rule based symbol manipulation) 이론-앞으로 고전주의 이론이라고 부르게 될-이 지배해 왔다. 근 사반세기의 지배였다. 더 정확히 말하자면 천구백사십년대에 이미 이 모형과 견줄 수 있는 모형, 곧 신경망 이론에 바탕을 둔 형식 모형-연결주의 모형-이 제시되었지만 그 모형이 갖는 개념상의 한계 때문에 그 동안 경쟁 상대가 되지 못했다. 고전주의는 '계산주의'란 다른 이름에서 알 수 있듯이 인간의 마음을 뉴엘(1980)이 부른 것처럼 물리적인 기호 체계로 본다. 마음이란 본질적으로 기호를 조작하는 정보 처리 시스템이고 그 처리 과정을 일종의 계산 과정으로 이해하고 있는 것이다. 디지털 컴퓨터 자체의 폭발적인 발전에 힘입어 규칙을 따르는 프로그램을 써서 문제를 해결한다는 모형은 모든 인지에도 적용되어서, 디지털 컴퓨터가 인간의 마음을 원리적으로 흉내낼 수 있다는 생각이 마음의 본성과 과정, 작용에 금급해 하는 인지과학자들의 관심을 끌게 되었다. 표상은 기호와 같다고 보았고 표상 상태는 디지털 컴퓨터의 내부 상태로 비유되었다. 처리 규칙은 기호적 표현을 조작하는 형식적 규칙과 같다고 보았고 그 규칙들은 컴퓨터의 프로그램 지시(instruction)들로 비유되었다.

* 이 글은 글쓴이의 석사 학위 논문(최 훈(1994a))을 줄이고 손 본 것이다. 석사 학위 논문의 초고는 "원정시합: 연결주의자와 구성적 구조"란 제목으로 1993년도 한국인지과학회 춘계학술발표대회(1993년 5월 8일 원광대학교)에서 발표한 것을 밝혀 둔다(한국인지과학회(펴냄), 1993, 『춘계학술발표대회논문집』, 47-57쪽).

포더가 생각하기에 이 모형은 '동네 안에 하나밖에 없는 놀이'였다 (Fodor, 1975, 27쪽, 55쪽). 그러나 이 전망은 너무 낙관적이었다. 하나밖에 없는 놀이는 우선 그 모형의 허물이 드러나기 시작했다. 사실은 연결주의가 등장하기 전까지 그것에 대한 반성이 전혀 없었다 해도 지나친 말이 아니다. 그런데 신경 생리학의 연구는 그 모형이 우리의 인지 현상과 들어맞지 않는 부분이 많다는 지적을 하고 있다. 여기에서 이마적에 연결주의(connectionism)라는 새로운 인지 모형이 재미있고 유익한 또 다른 놀이로 등장하기 시작했다.

상황이 이렇게 극적으로 변한 것은 친구백팔십년대 초반 이후의 일이다. 그 때를 전후해서 고전적인 패러다임과 뚜렷이 견주어지는 그림을 보여주는 듯한 인지 모형을 향하여 시선이 바뀌기 시작했다. 연결주의는 마음을 네트워크로 설명하려 한다. 이 네트워크는 서로 연결된 많은 수의 노드들로 이루어졌다. 이 노드들은 단순하며 그 수가 아주 많으며 연결주의 네트워크에 견주면 재래식이라 할 수 있는 폰 노이만식 컴퓨터와 달리 중앙 처리 장치의 통제를 받지 않는다. 연결주의에 마음이 끌리는 이들 중에는 이런 모형을 근거로 삼아 기호적 표상이 많은 인지적 작업에 필요하다는 사실을 부정하고 규칙이 명시적으로 있을 필요에 의심을 보냄으로써 고전주의라는 오쏘독스에 도전하는 이들도 있다. 이런 모형은 인지과학에 가히 혁명적인 반향을 불러일으켰으며 고전주의 이론에 쿨적인 위기를 몰고 왔다.

그러나 그 자리를 그렇게 쉽게 내줄 수는 없다. 고전적인 이론은 말한다. 연결주의 후보는 참신성과 실현 가능성을 동시에 가질 수 없다. 연결주의가 생물학적으로 더 실재적이라는 선행성을 바탕으로 인간이 하는 일, 특히 인지 활동에 더 가깝다는 공약을 내세우고 있지만 그 공약은 만약 실현 가능하다면 그리 새로운 게 아니다. 이미 우리 쪽에서 제시한 공약을 차원만 바꾸어서 실현한 것에 불과하다. 참신성이 없다. 또 만약 그 공약이 뭔가 새로운 게 있다면 그것은 실현 가능한 것이 아니다. 그야말로 빌 공자 공약이다.

물밑듯이 밀려오는 연결주의의 공격에서 이런 식으로 고전주의 이론을 지켜나가는 이들은 포더(Jerry Fodor)와 필리신(Zenon Pylyshyn)이다. 포더와 필리신의 논문 "연결주의와 인지적 구조: 그

비판적 분석”(1988)은 고전주의를 옹호한다는 단순한 수세적 입장에 그치지 않고 새로운 인지 모형이고자 하는 연결주의자를 딜레마에 빠뜨린다. 그 딜레마란 이런 식이다. 연결주의가 인지 체계에 대해 무언가 새로운 애기거리를 내세운다고 생각한다면 그 이론은 자신들이 인지 체계가 꼭 가져야 한다고 가정하는 어떤 본질적인 속성 X 를 가지지 못하는 한갓 연상주의(associationism)에 불과하다. 또 연결주의가 그런 X 까지 포섭할 수 있을 정도로 의미있는 이론이고자 한다면 그 이론은 더 이상 새로운 이론이 아니며 자신들이 옹호하는 고전주의 이론의 한갓 구현(implementation)에 불과하다.

이 논문은 연결주의에 대해 어떤 새로운 성과를 보태려고 하는 것은 아니다. 상당 부분이 위와 같은 주장을 담고 있는 포더와 필리신의 논문을 축으로 해서 진행될 것이다. 곧 포더와 필리신이 목표로 하는 바가 연결주의를 딜레마로 옹매으로써 연결주의의 도전으로부터 고전주의가 인지 이론으로서 갖는 지위를 보존하려는 것이었다면 글쓴이가 이 논문에서 하려고 맘먹은 바는 연결주의라는 새 이론의 도도한 흐름을 가로막는 포더와 필리신의 반발을 어루만짐으로써 그 흐름에 막힘이 없도록 하려는 것이다.¹⁾ 그런 목적을 위해서 이 글에서는 포더와 필리신이 내놓은 딜레마에서 첫번째 뿔을 다루려고 한다.²⁾

이 논문은 일단 연결주의 모형에서는 표상이 구성적 구조를 갖지 못한다는 주장에 대해서 집중하려고 한다. 우리의 논의 여행은 크게 보아 세 곳의 행선지를 거친다.

먼저 I 장에서는 포더와 필리신이 연결주의 모형의 표상에서는 드러나지 않는다고 한 구성적 구조가 도대체 무엇인가를 밝힌다. 일반적으로 구성적 구조를 자연 언어에 끌어들이는 논변은 생산성(또는

1) 이럴 수 있는 까닭은 과묵한 탓인지 모르지만 연결주의에 대한 철학적인 비판은 포더와 필리신 그리고 포더와 맥래플린의 논문이 지금까지는 거의 유일하고 다른 비판들도 그 논의를 벗어나지 못하고 있다는 믿음이 있기 때문이다. 사실 이 논문에서 포더와 그의 동조자들이라 부르는 이들은 이들 모두를 함께 가리킨다. 이런 상황에서 포더와 필리신을 극복한다는 것은 연결주의의 장래를 밝게 해줄 것이다.

2) 레마의 두번째 뿔, 곧 구현의 문제는 이 논문에서는 다루지 않는다. 구현의 문제에 대해서는 최 훈(1994b) 참조.

체계성) 논변이다. 곧 우리가 무한한 수의 아니면 한정없이 큰 수의 새로운 문장들을 만들 수 있고 이해할 수 있는 능력을 설명하려면 요소 표현이 있고 그 요소 표현으로부터 되부름적인 방식으로 복합 표현을 만든다는 식의 생각을 해야 한다. 대충 이러한 내용의 구성적 구조와 생산성 논변을 제시하고 구성적 구조에 최소한으로 요구된다고 글쓴이가 생각하는 조건을 제시한다. 그 조건은 (가) 요소, (나) 그 요소를 복합 표현에서 분해해 내고 다시 복합 표현으로 합성해 낼 수 있는 일반적이고 효율적이고 믿을 만한 방식의 두 가지이다. 포더와 필리신이 어떻게 해서 연결주의 모형에는 자신들이 생각하는 구성적 구조가 없다고 하는지를 알아 본다.

포더와 그의 동조자들이 구성적 구조를 근거로 연결주의를 비판하는 것에 대해 답변을 모색해 보는 것이 II장이 할 일이다. 가능한 응답으로서는 우리 표상이 구성적 구조를 필요로 한다는 전제는 인정하는 것이 있다. 그러나 이렇게 응답을 하는 이들은 포더와 그의 동조자들이 연결주의에 요구하는 구성적 구조는 너무 강한 구성적 구조이고 우리 표상에 필요한 구성적 구조는 약한 구성적 구조면 충분하다는 태도를 보인다. 따라서 이런 주장을 하는 이들의 목소리를 들어 보기 전에 구성적 구조의 약한 의미와 강한 의미가 어떻게 다른지 미리 알아 두어야 할 것이다. 남은 일은 연결주의가 약한 의미의 구성적 구조 정도는 보여 줄 수 있다는 증거를 대는 일이다. 이 일을 위해서 텐서 곱 표상 모형이 이 장의 마지막에서 제시될 것이다.

III장에서는 구성적 구조를 만족하기 위해 이 정도는 만족해야 한다고 생각했던 두 가지 조건 (가)와 (나)가 참된 구성적 구조가 되기 위한 필요 충분 조건임을 증명한다. 그래서 포더와 그의 동조자들이 마음에 두고 있는 강한 의미의 구성적 구조, 곧 고전적 요소성의 조건이 표상이 구성적 구조를 갖기 위해서 꼭 필요한 것이 아니라는 점을 밝히려고 한다. 그런 목적을 위해서는 그들이 왜 고전적 요소성이 구성적 구조에 필수적이라고 생각하는지 그 까닭을 들쳐내고 그 까닭이 근거없음을 지적하는 방식을 택할 것이다. 글쓴이가 생각하기에 그들이 고전적 요소성을 표상의 구성적 구조에 요구하는 까닭은 구체적으로 글자 그대로 사례화한 요소 표현만이 정신 현상에서 인과적인

힘을 발휘할 수 있다고 그들이 생각했다는 것이다. 이것에 대해 그 근거가 희박함을 지적한다. 이상은 표상이 구성적 구조를 갖추기 위해 (가)와 (나) 외에 요구하는 고전적 요소성 조건은 굳더더기 조건임을 주장하는 것인데 더 나아가 그 조건을 끝까지 주장하면 포더와 그의 동조자들이 일관된 주장을 하지 못하는 셈이 되어 모순에 빠지게 됨을 지적한다. 그래서 그들은 만약 약한 구성적 구조를 인정하면 연결주의도 구성적 구조를 지닌 표상을 가짐을 인정하는 셈이 되고 그렇지 않고 강한 구성적 구조를 고집하면 위와 같은 모순에 빠지게 된다는 새로운 딜레마에 처하게 된다.

포더와 필리신으로부터 시작된 연결주의 공격에서 위와 같은 구도를 가지고 연결주의를 성공적으로 지켜낸다면 연결주의는 적절하고 의미있는 인지 모형의 자격을 되찾기 위한 첫발을 내딛게 된다. 포더와 필리신이 옳맨 두 뿔 가운데 한쪽 뿔에서 빠져 나왔기 때문이다. 나머지 뿔마저 떨쳐버린다면 그들의 공격으로부터 완전히 벗어나는 길이 되겠지만 이 논문에서는 한 가지 작업에 만족하기로 한다.

I 구성적 구조

1.1 포더와 필리신의 연결주의 공격

포더와 필리신은 “연결주의와 인지적 구조: 그 비판적 분석”(1988)에서 새로운 인지 모형이고자 하는 연결주의자를 다음과 같은 딜레마에 빠뜨린다.

- (1) 인지 체계는 X를 가져야 한다. 따라서
- (2) 연결주의 체계는 (1) X를 가지지 못하거나—이 경우에 연결주의는 적합한 인지모형이 되기에는 너무 약하고, 아니면 (2) 연결주의 체계는 단지 고전적인 체계(기호 조작 모형)의 구현에 불과하다—이 경우에 연결주의는 인지에 대한 새로운 개념이 되지 못한다.

연결주의가 인지 체계에 대해 무언가 새로운 얘기거리를 내세운다고 생각한다면 그 이론은 자신들이 인지 체계가 꼭 가져야 한다고 가정하는 어떤 본질적인 속성 X 를 가지지 못하는 한갓 연상주의(associativism)에 불과하다. 또 연결주의가 그런 X 까지 포섭할 수 있을 정도로 의미있는 이론이고자 한다면 그 이론은 더 이상 새로운 이론이 아니며 자신들이 옹호하는 고전주의 이론의 한갓 구현(implementation)에 불과하다.

이 때 포더와 필리신이 X 로서 들고 있는 것은 바로 표상의 구성적 구조이다³⁾. 한 체계가 구성적 구조를 가졌다는 것은 그 체계에 기초가 되는 요소가 있고 그 요소들이 요소 표현들을 이루며 복합 표현은 그 요소 표현들로부터 되부름적(recursive) 규칙에 의해 결정된다는 것을 말한다. 곧 복잡한 표현의 내용은 그 복잡한 표현을 구성하고 있는(상대적으로 더) 간단한 부분의 내용에 의해 결정된다는 것이다. 포더와 필리신은 이런 구성적인 구조를 표상은 꼭 가지고 있어야 한다고 생각하는데 왜냐하면 그래야만 우리 표상에서 부인할 수 없는 현상인 생산성과 체계성을 설명할 수 있기 때문이다. 그런데 표상을 다루는 인지 모형이고자 하는 연결주의에서는 표상이 이 구성적 구조를 지니고 있지 못하다. 그러므로 연결주의는 적절한 인지 모형이 되기에는 약하다는 것이 포더와 필리신의 내세우는 바이다. 그런데 이런 구성적인 구조를 연결주의가 보여준다고 하더라도 그것이 고전적인 인지 구조를 구현하는 방법을 통해서라면 말짱 헛일이다. 그렇다면 그 이론은 인지 차원에서 주장되는 고전주의 인지 이론에 대한 물

3) 차머스는 표상의 구성적 구조가 표상이 가지는 본질적 속성 X 의 한 측면일 뿐임을 강조한다. 그래서 포더와 필리신이 구성성이 X 의 전부 다인 것처럼 생각하는 것은 잘못된 전제라고 말한다. 마음의 한 측면일 뿐인 그 구성적 구조를 연결주의는 못 보이고, 고전주의는 보였다고 해서 고전주의가 마음에 대한 최선의 이론이라는 결론을 내릴 수도 없으며, 연결주의가 구성성의 고전적 이론을 구현할 수 있다는 사실은 연결주의가 마음에 관한 고전적 이론을 구현한다는 것을 함축하지도 않는다고 한다(Chalmers, 1993). 차머스의 주장은 옳은 점도 있다. 그러나 그 주장이 포더와 필리신의 주장을 제대로 이해했느냐도 문제겠지만(Butler, 1993a, 참조), 구성적 구조가 마음이 가지는 한 측면일 뿐이라고 해도 중요한 측면임에 모두 동의한다면 그것을 설명해야 하는 의무를 연결주의가 벗어버리지는 못할 것이다.

리적인 구현 차원의 노드들이 될 뿐이므로 전혀 인지에 대한 고전주의 이론을 뒤집을 수 있는 새로운 이론이 되지 못하기 때문이다. 요컨대 연결주의 이론이 고전주의 이론의 대안이 되면서 적절한 이론이라면 고전주의 인지 구조의 한갓 구현이 되지 않으면서 구성적 구조를 설명할 수 있어야 한다는 것이 포더와 필리신의 주장인 것이다.

맥래플린을 따라 위 딜레마의 (a)와 (b)를 각각 적합성 조건과 비-구현 조건이라고 부르자(McLaughlin, 1993) 이 글은 포더와 필리신의 딜레마에서 첫째빨과 씨름하려는 의도였으므로 적합성 조건만 더 자세히 알아 보자.

1.2 적합성 조건⁴⁾

우리 자연 언어가, 더 엄격히 말하자면 자연 언어의 구문론과 의미론이 구성적이라는 것은 일반적으로 받아들여지고 있다.⁵⁾ 다시 말해서 자연 언어의 일반적인 구문론과 의미론은 기초가 되는 요소들과, 그 요소들로부터 이루어지는 더 복잡한 표현을 결정하는 (아마도 되부름적인) 규칙들로 이루어져 있는 것이다. 이를테면 ‘성자는 귀남이를 사랑한다’는 표현은 구성적인데 그 표현은 ‘성자’, ‘귀남’, ‘사랑한다’를 그 요소로 지니고 있고 그 요소들에 의해 ‘성자는 귀남이를 사랑한다’고 하는 복합 표현을 만든다. 이렇게 볼 때 구성성에서 요소(constituency)가 중요한 개념임을 알 수 있다. 복합 표현이 요소 표현으로 환원되고 이 요소 표현으로부터 복합 표현을 만들 수 있으며 그 과정이 되부름적인 규칙에 의해 일반적이고 믿을 만하고 효율적인 방법으로 이루어진다는 것이 구성성 개념의 중요한 축이다.

자연 언어가 구성적인 구문론과 의미론을 갖는다는 사실을 받아들일 수 있을까? 자연 언어에서 구성성을 원리로서 요구하는 동기는 무엇일까? 자연 언어에 구성적인 구조가 있다는 생각을 지지하는 근거

4) 최훈(1994a), 15-20쪽 참조.

5) 자연 언어의 구성적 성격을 부정하는 이들은 레이헤이브(1989)와 쉬퍼(1987, 1991)가 있다. 이들의 주장에 대해서는 최 훈(1994a), 2.2.1절과 2.2.2절 참조.

는 무엇일까? 이에 대해 두 가지 동기를 얻은 생각해 볼 수 있다. 첫째는 복합 표현을 요소 표현으로부터 말끔하게 설명할 수 있기 때문이다. 요소적 표현과 그 요소를 복합 표현으로 만들어 주는 구문론적 규칙이 그걸 가능하게 해준다. 또 다른 동기는 구성성이 자연 언어의 학습가능성(learnability)을 위해서 꼭 필요하기 때문이다(Davidson(1965) 참조). 이 점은 구성성에 대한 논의에서 보통 생산성 논변(productivity argument)이라고들 하는 것이다. 그 논변은 다음과 같다. 화자는 무한한 수의-아니면 적어도 한정없이 큰 수의-새로운 문장들을 이해할 수 있다. 그렇게 무한한 수의 문장을 이해한다는 것을 어떻게 설명해야 할까? 이에 대해서는 두 가지 대답이 가능하다. 첫째는 무한한 수의 문장 각각에 해당하는, 곧 그 문장이 뜻하는 무한한 수의 사실들이 있다고 가정하는 것이다. 이 무한한 사실들을 묶어 설명해 내는 유한한 토대가 되는 사실들은 있지 않다. 있는 것은 오직 무한한 사실들 뿐이다. 그러나 이 설명은 그럴 듯하지 못하다. 자연 언어를 쓰는 우리가 항상 새로운 문장에 부딪칠 때 그 문장을 이해해 내고 설명해 내며 또 새로운 문장을 생산해 내는 유한한 토대가 존재하지 않는다면 우리가 그 문장을 이해하고 학습한다는 것은 신비일 뿐이다. 설령 신비가 아니더라도 첫번째 동기에서 보이는 편리함과 말끔함을 지닐 수 없다. 그렇다면 우리는 다른 대답을 기다릴 수밖에 없다. 곧 우리가 무한히 많은 문장들을 이해할 수 있는 것은 우리가 문장을 이해한다는 것이 결국은 그 문장들의 부분들, 요소들을 이해하는 것이기 때문에 가능하다는 설명이다. 한 문장이 몇 개의 요소들로 이루어졌다고 하면 서로 다른 문장들은 구성 요소들의 종류와 이 요소들을 정렬하는 서로 다른 방식에 상응하게 된다. 그러면 새로운 문장은 새로운 정렬 방식과 상응하게 되는 것이다. 그러므로 한 문장의 의미는 그 요소 구조에 의해 결정된다. 이는 무한한 문장들이 주어졌을 때 좀더 유한한 요소들을 통해서 이해할 수 있고 또 이 요소들을 통해서 무한한 새로운 문장들을 만들어 낼 수 있다는 생각이다. 이렇게 봤을 때 생산성은 구성성에 의존하고 이 구성성에서는 요소성이 중요한 개념임을 알 수 있다. 포더의 지적처럼 “한 대상이 요소를 갖지 않는다면 그 대상의 의미를 요소의 의미로부

터 구성해낼 수 없다”(Fodor, 1987, 150쪽).

이상의 논변은 다음과 같은 구조를 갖는다.

- (1) 우리 자연 언어에는 무한히 많은 문장들이 있고 우리는 그 문장들을 이해한다.
- (2) 우리 자연 언어가 구성적인 구조를 가지면 (1)이 잘 설명된다.
- (3) 그러므로 우리 자연 언어는 구성적인 구조를 갖는다.

(1)은 우리 언어의 생산성이라는 경험적 사실이다. 구성성은 그 경험적 사실을 설명하기 위해 요청되는 원리이다. 우리 언어에 구성적 구조를 가정하면 언어의 생산성이라는 경험적 사실을 잘 설명할 수 있다는 게 이 논변의 구조이다.

이 구조에 대해서 두 가지 점을 지적해야겠다. 첫째, 이 구조는 바로 ‘최선의 설명으로 만드는 추론’(inference to the best explanation)의 형태를 띠고 있다. 그렇다면 이런 논변은 같은 경험적 현상에 대해 의미있는 대안이 되는 설명이 더 있다는 것을 제거하지 못하면 후견 긍정의 오류를 범하게 된다. 쉬퍼(Stephen Schiffer)가 자연 언어의 구성적 구조를 가정하지 않아도 생산성을 설명해 낼 수 있다고 하여 구성적 의미론을 부정하는 것도 바로 이 대목에서다.⁶⁾ 둘째, 포더와 필리신은 자연 언어의 구성적 구조를 받아들이게 하는 근거가 되는 (1)의 생산성이라는 현상이 너무 강한 주장이어서 그걸 받아들이지 않는 이들이 있을까 조심스러워한다. 그래서 우리 자연 언어에서 발견할 수 있는 다른 현상, 그러면서 어느 누구나 동의하리라고 생각하는 현상, 곧 체계성(systematicity)도 구성적 구조를 지지할 수 있는 근거라고 내세운다. 언어 능력이 ‘체계적’이라는 말로 뜻하는 바는 어떤 문장을 말하거나 이해하는 능력은 다른 많은 관련된 문장들을 말하거나 이해하는 능력과 본질적으로 연관되어 있다는 것이다(Fodor/Pylyshyn, 1988, 37쪽; Fodor, 1987, 149쪽). 이를테면 “성자는 귀남이를 사랑한다”를 말하고 이해할 줄 아는 이가 “귀남이는 성자를 사랑한다”를 말하고 이해할 줄 모를 리가 없다. 이 체계성

6) 이에 대해서는 최 훈(1994a), 2.2.2절 참조.

도 우리 언어가 구성적인 구조를 가졌기에 설명할 수 있다. “성자는 귀남이를 사랑한다”는 문장을 ‘성자’, ‘귀남’, ‘사랑한다’의 요소로 분해할 수 있고 그 요소를 통해 다시 “귀남이는 성자를 사랑한다”를 만들 수 있기에 “귀남이는 성자를 사랑한다”는 문장을 이해할 수 있는 것이다. 결국 체계성도 구성성을 요구하게 되고 구성성에서는 요소라는 개념이 중요한 구실을 함을 다시 한 번 확인할 수 있다.

지금까지의 논의를 그림으로 정리해 보자.

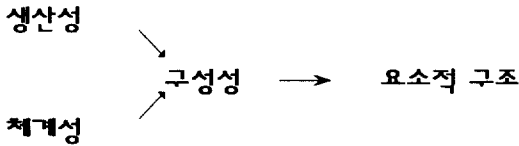


그림. 생산성 논변 ($P \rightarrow Q$ 는 P 를 설명하는 데 Q 가 요구된다는 뜻)

여기서 중요한 점은 구성성을 지지하는 근거가 언어의 생산성인가 체계성인가 하는 것이 아니다. 오히려 두 가지 현상 모두가 구성성을 떠받들어 주는 경험적 현상이라고 생각할 수 있다. 그리고 구성성이 두 현상을 설명할 수 있는 까닭은 요소적인 구조를 지니고 있기 때문이다. 그런데 그렇다고 해서 생산성과 체계성을 설명하는 최종적인 근거가 요소적인 구조라고 생각해서는 안된다. 한 복합 표현이 요소 표현을 갖는다는 점만으로는 구성적 구조를 다 말할 수 없기 때문이다. 그냥 요소를 갖는 것만으로는 부족하고 그 요소에 민감한 구조여야 한다. 다시 말해서 그 요소로부터 인과적인 영향을 받아야 한다. 요소가 그냥 요소여서는 안되고 복합 표현이 그 요소로부터 인과적인 영향을 받을 만한 요소가 되려면 무슨 조건이 거기에 보충되어야 할까? 글쓴이는 요소가 구성적 구조를 설명해 낼 수 있는 참된 요소가 되기 위해 보충되어야 할 조건을 두 가지 들겠다. 하나는 요소가 의미있는 단위여야 한다는 조건이고 다른 하나는 요소 표현이 그 요소가 나타나는 어느 복합 요소에서나 한결같이 적용되어야 한다는 조건이다.

첫번째로는 요소가 복합 표현에서 아무렇게나 잘게 나누어 놓은 단위가 되어서는 안되고 그 표현이 이루어지는 체계 안에서 충분한 의미를 가질 수 있는 단위가 되어야 한다는 사실이다. 보기를 들어보

면 “성자는 귀남이를 사랑한다”가 ‘성자’, ‘는귀남이’, ‘을사’, ‘랑한다’와 같은 요소를 갖는다는 것은 언어의 생산성과 체계성을 설명하는데 아무런 도움도 못된다. 그 요소로는 분해와 합성이 일관되지 못하다. 그 요소를 가지고 “귀남이는 성자를 사랑한다”를 합성해 낸 다음에 다시 그런 요소들을 분해해 낼 수 있다고 생각할 만한 일반적이고 효율적이고 믿을 만한 방법은 없기 때문이다. 사람 한 명 한 명을 의미있는 단위로 생각하고 논의하는 체계를 생각해 보자. 그 체계에서 사람 한 명 또는 몇 명이 모인 집단은 의미있는 논의 단위가 될 수 있겠지만 사람의 어느 신체 부위는 의미있는 단위로 생각할 수 없을 것이다. 이 때 한 체계에서 어느 정도의 단위까지가 의미있는 단위인가 하는 물음이 생길 수 있을 텐데 그것은 그 체계가 처한 상황이 결정할 문제이다. 두 번째 조건은 요소의 한결같은 적용 가능성 조건⁷⁾으로서 모든 요소들은 그 요소가 나타나는 모든 정상적인⁸⁾ 복합 표현에서 한결같이 적용되어야 한다는 것이다. ‘성자’라는 요소가 “성자는 귀남이를 사랑한다”는 복합 표현에서 갖는 의미와 기능이 “성자는 귀남이를 미워한다”에서 갖는 의미와 기능과 서로 다르다면 애초에 구성성으로부터 우리가 원하는 바인 언어의 생산성 또는 체계성을 얻을 수가 없다. 그 두 문장은 ‘성자’, ‘귀남’, ‘사랑한다’ 등 똑같은 부분들로부터 만들어야만 한다.”

이상의 두 조건은 결국은 반 켈더가 말했듯이 한 복합 표현이 요소

-
- 7) 이 조건은 레이헤브가 말한 형용사의 적용 가능성 조건(applicability condition)을 모든 요소에 일반화한 것이다. 그러나 그 의미는 약간 다르다. 일반적인 적용 가능성 조건에 대해서는 2.2.1절과 Lahav(1989), 262쪽 참조. 그리고 한결같은 적용 가능성 조건과 일반적인 적용 가능성 조건의 차이점에 대해서는 최 훈(1994a), 2.2.3절 참조.
- 8) 은유적인 표현의 경우에는 예외를 인정한다는 말이다.
- 9) 이 조건을 레이헤브식의 일반적인 적용 가능성 조건으로 보면 개념의 고정성(stability) 또는 문맥 독립성을 요구하는 조건으로서 연결주의의 주장과 정면으로 맞붙는 뜨거운 문제가 될 수 있다. 포더와 그의 동조자들은 일관되게 한 개념이 어느 문맥에서나 같은 의미를 지녀야 함을 주장하나(Fodor/Pylyshyn, 1988, 20쪽; Fodor/McLaughlin, 1991, 339-340쪽), 연결주의는 개념의 문맥 민감성(Goschke/Koppelberg, 1990, 259-61쪽), 근사한 개념과 가족 유사성(Smolensky, 1991, 208-10쪽) 등의 표현을 통해 그 것을 부정한다. 최 훈(1994a), I 장 4절, 2.2.1절, 2.3.3절을 보라.

표현을 가질 때 아무 원칙이 없이 이루어지는 것이 아니라, 그 분해를 할 때 그리고 그 분해한 요소를 다시 복합 표현으로 만들 때 '일반적이고 효율적이고 믿을 만한' 방식이 있는 한해서 요소적 구조가 구성성에서 의미를 가져야 한다는 조건을 따르는 것이 된다(van Gelder, 1990, 361쪽 참조). 이 때 '일반적'이란 어떠한 임의의 복합적인 표현에도 분해와 합성의 방식을 적용할 수 있다는 뜻이고, '효율적'이란 그 방식이 기계적으로 구현이 가능하다는 뜻이고, '믿을 만하다'는 것은 같은 입력값에 대해서는 언제나 같은 산출값을 낸다는 뜻이다. 위에서 말한 두번째 조건에서 같은 요소가 어느 복합 요소에서나 한결같이 적용되어야 한다고 말하는 것은 같은 입력값에 대해서 언제나 같은 산출값을 내야한다는 믿을 만한 방식을 나타낸다. 또 첫번째 조건에서 의미있는 단위를 상정해야 한다고 말하는 것은 일반적인 분해, 합성 방식을 위한 전제 조건이 된다. 또 그렇게 해야만 각 표현에 대해 주먹구구식의 방식이 아닌 효율적인 분해, 합성 방식이 가능해진다.

이상과 같은 두 조건은 명시적인 설명이 뒤따라지는 않지만 포더와 필리신이 다음과 같은 귀절에서 주장하는 바다.

(분자)표상의 의미론적 내용은...그 표상의 구문론적 부분이 갖는 의미론적 내용의 함수이다(Fodor/Pylyshyn, 1988, 12쪽).

이 귀절에서 포더와 필리신이 함수란 표현을 통해 노리는 바는 두 가지인 것 같다. 하나는 함수의 출력값은 항상 입력값에 의해 결정된다는 사실이고 다른 하나는 같은 입력값에 대해 항상 일정한 출력값을 내놓는다는 사실이다. 이렇게 봤을 때 포더와 필리신도 구성성을 얘기하면서 드러내놓고 말하지는 않았지만 복합 표현은 요소 표현을 가질 뿐만 아니라 그 복합 표현의 의미가 요소 표현의 의미에 의해 결정되고 그 결정 과정 그리고 그 요소들에 대한 분해와 합성 과정이 아무 원칙이 없이 되는 것이 아니라 일반적이고 효율적이며 믿을 만한 방식을 따른다는 데 동의할 것이다.

이상의 논의를 통해서 그림 1을 읽을 때 생산성 또는 체계성을 설명해 내는 근거는 구성성으로 보아야 하고 요소는 구성성을 이루는

중요한 한 성분임이라는 데 주의를 하면서 구성성과 요소적 구조 사이에는 위와 같은 긴장 관계가 있다는 것을 알아야 할 것이다. 그렇다면 정리해서 구성적 구조는 다음과 같은 조건을 만족하는 구조로 보아야 한다.

(가) 요소

(나) 그 요소를 복합 표현에서 분해해 내고 다시 복합 표현으로 합성해 낼 수 있는 일반적이고 효율적이고 믿을 만한 방식

(가)와 (나)는 서로를 제약한다. (가)의 요소가 있어야 조건 (나)가 의미가 있다는 것은 사소하다. 그리고 (나)를 통해서 (가)의 요소가 위에서 말한 두 조건을 만족하는 참된 요소가 된다. 이 두 조건을 구성적 구조의 조건이라고 부른다.¹⁰⁾

이 조건은, III장에서 주장하겠지만, 구성적 구조의 필요 충분 조건이다. 필요 충분 조건임을 주장하는 전략은 다음과 같다. 자연 언어의 구성적 구조가 왜 필요했는가? 그것을 되새겨 보라. 바로 자연 언어의 생산성, 체계성을 설명하기 위해서 구성적 구조를 끌어들었다. 그렇다면 우리에게 필요한 구성적 구조는 그 생산성과 체계성 현상을 올바르게 설명할 수 있는 조건을 갖추면 충분하다. 그리고 이 조건만 만족하면, 곧 II장에서 약한 구성성 또는 기능적인 구성성으로 부를 것을 만족하면 생산성, 체계성 현상을 설명하는 데 필요한 것을 모두 만족한다. 그 외의 것은 굳더더기이다. 한 마디로 말해 이 조건을 통해서 자연 언어의 구성적 구조가 요구되는 상황을 충분히 설명해 낼 수 있고 이 조건 외의 다른 조건은 필요치 않다는 것을 주장하겠다.¹¹⁾ 이러한 연유로 우리는 구성성 개념이 도입되는 생산성 논변을 잘 기억하고 있어야 할 것이다.

포더와 펠리신이 보기에 연결주의와 고전주의가 차이를 보이는 것

10) II장 1절에서는 이 두 조건이 '약한 의미의' 구성적 구조의 두 조건이 된다.

11) (나)의 조건을 명시적으로 말한 반 겔더의 논문도 이 과정은 다루지 않고 있다. 만약 글쓴이의 이 과정이 성공한다면 그것은 반 겔더의 표현을 빌린 글쓴이의 논문이 갖는 독창성이 될 것이다.

은 그 이론들 사이에 구조적인 차이가 두 가지 있기 때문이다 (Fodor/ Pylyshyn, 1988, 12-13쪽).

(1) 정신적인 표상에 관한 구성적인 구문론과 의미론. 사고는 구성적인 구조를 갖는다.

이 말의 뜻은 I 장 2절에서 잠깐 말한 것과 같다. 곧 '성자는 귀남이를 사랑한다'는 더 이상의 요소를 갖지 않는 사고가 아니다. 다시 말해서 원자적이지 아니다. '성자'와 '귀남'에 대한 요소 사고로부터 만들어진 구성적인 정신 상태이다.

(2) 과정의 구조 민감성. 정신 과정(mental process)은 구문론적인 구성적 구조에 민감하다¹²⁾.

이들테면, P & Q의 꼴을 갖는 어떠한 사고에서도 P와 Q가 무었이나에 상관없이 우리는 P를 이끌어 낼 수 있다.

포더와 필리신은 (1)과 (2)를 고전주의를 정의하는 것으로까지 격상시키고¹³⁾ 그리고 나서 연결주의를 분석한다. 포더와 필리신은 연결주의에서는 모든 표상이 원자적이라고 주장한다. 곧 요소를 갖는 복합적인 표상이란 게 없다. 정신 상태는 구성적인 구조를 갖지 못하고 그러기에 (1)을 위배한다. 더구나, 연결주의의 과정은 구문론적인 구조에 민감하지 못하고 통계적인 속성, 곧 빈도(frequency)에만 민감하다. 이점은 전통적인 연상주의와 유사한데 그러기에 (2)를 위배하고 있다(Fodor/Pylyshyn, 1988, 31쪽).

그런데 포더와 필리신이 연결주의를 공격하는 잣대로 쓰는 구성적

12) '민감하다'(sensitive)는 말은 '인과적으로 영향을 끼친다' 정도의 뜻일 게다.

13) 반겔더는 이게 포더와 필리신뿐만 아니라 모든 고전주의자한테 최소한의 두 가지 주장임을 인정한다(van Gelder, 1991a, 357쪽). 만일 그렇다면 어떤 인지 모형이 고전주의와 양립 가능한가를 확인해 보려면 그 모형이 이 두 가지 주장을 만족하는가를 알아보면 될 것이다. 이 논문에 깔려 있는 생각은 연결주의 모형이 그 두 가지 주장을 만족할 뿐만 아니라 고전주의가 약점을 보이는 곳에서 더 나은 설명을 하는 세련된(refined) 인지 모형이라는 것이다.

구조, 다시 말해서 그들이 생각하는 구성적 구조도 글쓴이가 참된 구성적 구조의 조건이라고 말한 것을 만족할까? 아니면 그것을 넘어서거나 못미치는 것일까? 이에 대해서는 II장 1절에서 말하고 그게 정당한 것인지는 III장에서 따지겠다.¹⁴⁾

1.3 구성적 구조가 없는 연결주의 모형

자연 언어의, 그리고 표상의 구성적 구조를 앞 절에서와 같이 이해했다고 할 때 포더와 필리신은 연결주의 모형이 그런 구성적 구조를 보이지 못한다고 주장한다.

연결주의는 유닛들 사이의 기본적인 관계로서 인과적인 관련성만을 인정한다. 만약 당신이 유닛들 사이에 흥분과 억제가 어떻게 일어나는가 알고자 할 때 네트워크 안에 있는 유닛들이 서로 어떻게 관련을 맺는가에 대해서만 알면 다 된다. 이에 비해 고전적인 이론은 의미론적인 가치를 부여한 대상들의 인과적인 관계뿐만 아니라 구성성이 성립되는 구조적인 관계까지도 인정한다(Fodor / Pylyshyn, 1988, 12쪽).

결론은 간단하다. 자신들이 주장해온 고전주의 모형은 구성적인 구조를 보이고 있지만 연결주의는 구성적 구조를 보이지 못한다는 것이

14) 한가지 재미있는 사실은 포더와 필리신은 위에서 인용한 그들 논문 12쪽의 귀절 말고 정신 표상의 구성적인 구문론과 의미론이 만족해야 하는 성질을 두 가지 더 말한다. 그것은 각각 “구조적으로 원자적인 표상과 구조적으로 분자적인 표상 사이에 구분이 있다.”는 것과 “구조적으로 분자적인 표상들은 구문론적인 요소들을 갖는데 그 요소는 구조적으로 분자적이거나 구조적으로 원자적이다.”는 것이다. 이 두 성질은 ‘요소’의 성격을 설명한 것이다. 그렇다면 그들은 분명 처음에는 “정신 표상의 구성적 구문론과 의미론”이란 제목을 통해, 글쓴이가 말한 (가)와 (나)의 조건 이상의 것을 말하지 않고 있다. 그런데 그 이후 여러 비판(최 훈(1994a), II장 참조)에 직면한 포더는 맥래플린과 손잡은 논문에서는 글쓴이가 III장에서 말할 (다)라는 조건을 더 보태는 것이다. 물론 그렇다고 해서 포더와 필리신의 논문이 그 조건에 대해 언급을 전혀 안하는 것은 아니다. 다만 포더와 맥래플린의 논문에 견주어 봤을 때 표현의 명시도나 강조의 정도에서 덜하다는 것이다. 덜하다는 것은 다시 말하면 그 때만 해도 조건 (다)가 필요함을 심각하게 느끼지 못했다는 말이 될 것이다.

다. 그게 두 이론 사이의 차이점이고 그 차이점은 연결주의로부터 의미있는 인지 이론의 지위를 뺏을 수 있을 만큼 큰 것이다. 연결주의는 왜 구성적 구조를 보이지 못한다고 생각하는가? 연결주의 체계에서는 개별적인 유닛이나 유닛의 연합이 의미론적으로 해석될 수 있다. 그러나 유닛들 사이의 기본적인 관계는 인과적인 관계만을 인정한다. 이 인과적인 관계에 대해서 포더와 필리신은 자세한 설명을 하고 있지 않으나 짐작컨데 유닛과 유닛 사이에 연상적인 관계만을 인정하고 부분과 전체의 관계일 때 나올 수 있는 포함과 도출 관계는 전혀 고려하지 않는 관계이다. 있는 것이라고는 오직 유닛 사이에 미치는 힘-고전적인 연상주의에서는 빈도, 연결주의에서는 활성화 정도-뿐이다. 모든 유닛은 각각이 지니는 그리고 서로에 미치는 힘을 제외하고는 구분이 되지 않는 원자적인 상태이다. 또 구문론적 규칙을 적용하려면 적당한 요소가 있어서 거기에 되부름적인 규칙을 쓸텐데 유닛 하나하나가 모두 원자적인 상태이니 구문론적 규칙에 따라 조작할 나위도 없다(최 훈, 1994a, 30-3쪽 참조).

이와 같은 포더와 필리신의 비판에 대해 연결주의자의 처지에서는 개념적인 답변과 경험적인 답변이 가능하다. 여기서는 경험적인 답변, 그 중에서도 특히 스몰렌스키의 텐서 곱 표상을 포더와 필리신 논변에 대한 반례로서 내놓으려고 한다.¹⁵⁾ 경험적인 답변이 철학자들에게는 어떻게 보면 더 강력해 보이고, 또 포더와 맥래플린이 이 반례에 대한 해석을 놓고 맞대응을 하는 것이 다음 답변의 다음 절과 그 다음 절의 논의거리가 되기 때문이다.

II 연결주의자의 답변

연결주의 모형이 구성적 구조를 못 보인다는 포더와 필리신의 논변에 대해 지금까지는 표상이 구성적 구조를 갖는다는 사실을 애초에 부정하여 연결주의를 포더와 그의 동조자들의 공격으로부터 막으려는

15) 개념적인 답변에 대해서는 최 훈(1994a), II장 2절과 3절 참조.

논변이 있을 수 있다. 그런데 여기서는 표상이 구성적 구조를 갖는다는 사실을 인정하면서 그들과 맞서는 논변을 살펴보도록 하자.¹⁶⁾ 곧 연결주의자도 표상은 생산성이나 체계성을 설명하기 위해서는 구성적 구조를 가질 수밖에 없다고 생각하며, 그 이론이 표상의 의미있는 인지 이론이고자 한다면 표상의 그런 성격을 보여줄 수 있어야 하며 연결주의는 그럴 수 있는 모형이라는 논변이다. 그러나 이 때 연결주의의 표상이 구성적 구조를 갖느냐는, 사실의 확인을 요구하는 사안에 대해 이들과 포더 진영의 생각이 달라지는 것은 아니다. 그것보다는 그 이전에 이미 구성적 구조의 정체에 대해 지금부터 보려는 주장과 포더와 그의 동조자들의 의견이 갈라서는 것이다. 결론부터 말하면 포더와 그의 동조자들이 표상에 요구하는 구성성은 강한 구성성이고 연결주의가 보일 수 있는 구성성은 약한 구성성이다. 그런데 정작 우리 표상에 필요한 것은 약한 구성적 구조이다. 이 절 전체의 논의를 위해서 가장 먼저 앞에서 했던 논의와 연관지어 약한 구성성과 강한 구성성을 다시 구분해 보겠다(2.1절).

2.1. 구성적 구조의 조건: 약한 구성성과 강한 구성성

표상의 (강한) 언어성과 (약한) 언어성을 구분해 보자. (강한) 언어성은 (약한) 언어성과 달리 언어와 같은 기호 체계가 (가)와 (나)의 조건 외에 다음과 같은 의미를 더 갖는다고 생각하는 언어성이다.

(3) 단순 기호와 복합 기호 사이에는 구체적인 의미에서 부분-전체 관계가 있고 단순 기호는 그 단순 기호가 나타나는 어떤 복합 기호에서도 그대로 간직된다.

그런데 덧붙인 이 의미는 포더와 그의 동조자들이 표상이 갖추어야 하는 조건으로 분명하게 밝히고 있는 고전적인 요소성(classical constituency)이라는 개념이 표현하고 있는 것이다. 곧 한 쌍의 표현 E1과 E2에 대해서 E1이 E2의 고전적 요소인 것은 E2가 사례화할

16) 앞논변에 대해서는 최 훈(1994a), II장 2절 참조.

때마다 E1이 사례화할 때 뿐이다. 이 고전적인 요소성을 요구하는 구성성이 바로 강한 의미의 구성성이고 포더와 그의 동조자들이 표상이 구성적 구조를 가져야 한다고 할 때 생각하는 구성적 구조이다. 강한 의미의 구성적 구조라고 부르는 것은 그 자체가 글자 그대로(literal) 나타나는 부분-전체 관계에서 직접적으로 보이는 구조를 말한다. 표현의 유형(type)과 사례(token)를 구분한다고 할 때 복합 표현의 사례가 실제로 그 표현의 요소 표현의 사례를 포함하는 경우를 말한다. 곧 이를테면 ((P & Q) & R)이라는 표현이 (P & Q)라는 사례화된 표현을 아주 구체적으로, 추상적인 의미에서가 아닌 방식으로 포함하는 경우에, 또 I장 4절에서 쓴 용어로 말해 보면 그 요소 표현이 구현될 때 우리는 강한 의미의 구성적 구조라고 말한다. 이 의미에서 복합 표현을 요소 표현으로부터 구성할 때 거꾸로 복합 표현을 요소 표현으로 분해할 때 일단 문법적인 규칙에 따를테고 중요한 것은 구성 요소를 그대로 간직한다는 점이다.¹⁷⁾

그러나 복합 표현이 요소 표현으로부터 글자 그대로의 방법에 의해 구성되지 않을 수도 있다. 곧 약한 의미의 구성적 구조의 가능성이 있다. I장 2절에서 구성적 구조의 조건을 말했던 것을 되새겨 보자. 거기서 성급하게 구성성의 필요 충분 조건이라고 했는데 그 구성적 구조는 다음과 같은 두 가지 조건을 만족하는 구조였다.

(가) 요소

(나) 그 요소를 복합 표현에서 분해해 내고 다시 합성해 낼 수 있는 일반적이고 효율적이고 믿을 만한 방식

강한 구성성도 이 두 조건을 만족하기는 한다. 그러나 강한 구성성의 글자 그대로의 구문론적인 방식이 아닌 다른 방식으로, 이 두 가지 조건만을 만족하는 구성성을 약한 의미의 구성성이라고 부르자.

17) 반젤더는 강한 의미의 구성적 구조에 '연쇄적'(concatenative)이라는 낱말을 붙인다(van Gelder, 1990, 1991a; 이 논문의 2.3.4절 참조). 또 강한 의미의 구성성은 그 설명에 비추어 볼 때 '구문론적 구성성'이라고 불려도 좋을 것이다. 표현 유형들 사이에 나타나는 추상적인 요소 관계가 그와 상응하는 사례들의 물리적 구조에서 직접적이고 구체적으로 사례화된다는 의미에서 말이다.

고쉬케와 코펠버그 그리고 반 켈더가 명시적으로 또 클락이 간접적으로 연결주의 네트워크가 구성적인 구조를 갖는다고 말할 때가 바로 이런 경우인데, 구성 요소가 되는 사례를 그 사례가 포함되어 있는 복합 표현에서도 단순히 그대로 간직하는 방법을 쓰지 않고서도 일반적이고, 효율적이고, 믿을 만한 구성과 분해의 방법을 통해서 표상이 구성적인 구조를 갖는다고 말하는 것이 적절할 수 있다. 청동은 구리와 주석의 합성물이다. 그러나 청동의 구체적 사례들 속에는 그것의 요소들인 구리와 주석이 구체적으로 사례화되지는 않는다. 그렇지만 우리는 청동에서 그 요소들인 구리와 주석을 일반적이고 효율적이고 믿을 만한 화학적인 방법으로 추출해 낼 수 있고 또 거꾸로 구리와 주석으로부터 청동을 합성할 수 있다. 우리 자연에는 이런 합성물들이 얼마든지 있다.

(가)와 (나) 조건만을 만족하는 약한 의미의 구성성을 직관적으로 이해하기 위해서 반 켈더가 든 어둠상자(black box)의 비유를 생각해 보자(van Gelder, 1991a, 364쪽). 먼저 (강한 의미의) 구성적인 성격을 띤 표현(이를테면 영어문장)과 벡터 공간 속의 원소들 사이에 상응 관계가 있다고 가정하자. 그리고 어둠상자가 두 개 있다고 상상해 보자. 이 상자들은 보통의 기계 장치이다. 그리고 그 상자를 어둠상자라고 부르는 까닭은 딱 게 아니라 지금은 그 상자가 어떻게 작동하는지에 관심이 없기 때문이다. 이 상자 가운데 하나는 구성, 곧 합성 장치이다. 그것은 한쪽 끝에서 한 복합 표현의 요소 표현들에 해당하는 벡터들을 주면 다른 쪽 끝에서 그 복합 표현에 해당하는 벡터를 뱉어 낸다. 다른 상자는 이 동작과 반대로 작동하는 분해 상자이다. 곧 한쪽 끝에서 복합 벡터를 주면 그 상자는 그걸 잘 ‘뺏아서’ 다른 쪽 끝으로 요소들의 집합을 뱉어낸다. 여기서 벡터 공간의 원소들을 표상으로 생각한다면 우리는 분명히 요소를 가지게 되는 것이다. 그런데 이 때 요소란 강한 의미의 구성적인 구조에서 말하는 종류의 글자 그대로의 부분인 요소가 아니라 단지 기능적인 부분일 뿐인 요소이다.¹⁸⁾

18) 반켈더가 어둠상자로써 비유를 든 것은 적절한 것 같다. 왜냐하면 전기 회로나 기계, 생물학적인 체계 등에 대해서 그 내부 구조는 문체삼지 않고 그 기능 또는 그것에 대한 입력과 출력의 관계만을 고찰의 대상으로 여기는 과정 또는 그

이 때 반 켈더 스스로는 드러내놓고 강조하지 않았지만 이 비유가 약한 의미의 구성성을 제대로 드러내기 위해서는 지켜야 할 게 더 있다. 구성적 구조의 조건을 내놓을 때도 말했듯이 단순히 요소만 가진다는 것으로 그쳐서는 안된다. 그것은 구성적 구조의 조건 (가)와 (나) 두 조건 가운데 하나만 만족하는 것이다. 그러므로 그 요소는 같은 복합 벡터에서 분해될 때 어느 때라도 같은 결과를 갖는 요소여야 한다. 또 그 요소를 분해할 때 그 같은 요소를 산출해 내는 효율적인 방법이 있어야 한다. 한편 그 반대의 경우도 마찬가지이다. 곧 요소에 해당하는 벡터를 집어 넣어 복합 표현에 해당하는 벡터를 뱉어낼 때 마찬가지로 어느 때라도 같은 입력에는 같은 출력을 얻을 수 있어야 하며 출력을 얻는 방법도 효율적이어야 한다.

이상의 설명으로 약한 구성성과 강한 구성성이 어떻게 다른지 그 차이점이 드러났을 것이다. 확인하자면 강한 구성성에서는 복합 표현이 요소 표현을 항상 글자 그대로 포함하고 요소 표현은 분해와 합성의 과정에서 그 형태를 그대로 간직해야 한다. 이 점을 드러내 줄 수 있는 조건을 앞에서 말한 구성적 구조의 조건 (가)와 (나)에 덧붙여 조건 (다)로 제시해 보자.

(다) 한 표현이 다른 표현의 요소일 수 있는 경우는 다른 표현이 구현될 때마다 그 표현이 구현되는 경우일 때만이다.

이것은 위에서 말한 (강한) 언어성을 띤 기호 체계에 덧붙인 의미 (3)을 다시 쓴 것에 지나지 않긴 하다. 보기를 들어보면 “성자”는 “성자는 귀남이를 사랑한다”의 요소가 되고 그러기에 두 표현은 온당한 구성적 구조를 보여주고 있는데 왜냐하면 조건 (다)가 말하듯이 “성자는 귀남이를 사랑한다”는 표현이 나타날 때마다 “성자”라는 표현이 나타나기 때문이다. 그런데 주의할 것은 이 조건은 연결주의자 편에서 고전주의자의 구성성 주장을 감당하기 힘든 것으로 만들기 위하여 덧붙인 것이 아니라는 점이다. 조건 (다)는 이미 보았듯이 바로 포더

와 그의 동조자들이 ‘고전적 요소’라는 이름으로 불리는 조건을 규정 한 것이다(Fodor/McLaughlin, 333쪽). 그리고 I 장 3절에서 포더와 필리신이 연결주의 네트워크에는 구성적 구조가 없다고 몰아붙이는 주장을 잘 들여다보면 이 조건을 발견할 수 있을 것이다.

이렇게 볼 때 우리는 강한 의미의 구성적인 구조로 구현된 종류가 아닌 구성적인 구조를 가질 수 있음을 대강이나마 이해하게 되었다. 그렇다면 이제 할 일은 분명하다. 두 가지 할 일이 우리 앞에 던져져 있다. 우선은 두 가지 구성적 구조 가운데 어느 것이 참된 구성적 구조인지를 밝히는 것이다.¹⁹⁾ 연결주의 모형도 구성적인 구조를 가지고 있음을 보여주려는 게 이 논문의 목표였으므로 그 목표를 달성하기 위해서는 약한 구성성이 충분하고도 필요한 구성성임을 보여야 될 것이다. 그것을 보일 수 있는 방법은 우리 자연 언어 나아가 사고에 구성적 의미론이 애초에 요구되는 상황, 곧 생산성 논변을 생각해 볼 때 그 요구를 충족시키기 위해서는 얼마나 강한 구성적 구조가 필요한가에 초점을 맞추면 된다. 글쓴이는 더 나아가 III장에서 (가)와 (나) 조건이 참된 구성적 조건이고 (다)는 군더더기인 조건임을, 나아가서는 그걸 가정함이 모순을 불러일으킴을 주장하겠다. 둘째로 그 일이 완성된다고 하더라도 잊지 말아야 할 일은 연결주의 체계에서 약한 구성성을 보여주는 표상이 있음을 보여주는 일이다. 그것에 이어 약한 의미의 구성적 구조를 본격적으로 보여준다고 생각하는 스몰렌스키의 텐서 곱 표상을 검토하겠다(2.2절). 그리고 나서 참된 구성적 구조를 밝히는 작업을 III장에서 하겠다.

2.2. 스몰렌스키의 텐서곱 표상

스몰렌스키의 근본적인 생각은 구성적 구조를 벡터 연산으로 표상하려는 것이다. 그 과정을 Smolensky(1990)에 제시된 생각을 따라

19) 구성성이 약한 의미와 강한 의미로 나누어진다는 것은 그리 새로운 생각이 아니다. 그러나 그런 생각을 펼치는 이들, 반 켈더나 타이, 스몰렌스키도 어느 정도의 구성성이 참된 구성성인지를 적극적으로 밝히지 않고 있다. 물론 그 반대편에서 있는 포더와 그의 동조자들은 강한 의미의 구성성이 참된다는 주장을 적극적으로 하고 있지만 말이다.

가면서 “성자는 귀남이를 사랑한다”고 하는 구성적 구조를 어떻게 연결주의 네트워크가 표현해 내는지 알아보자.

먼저 익혀야 할 개념들이 있다. 스몰렌스키의 텐서 곱 표상에서는 구조화된, 곧 구성적인 구조를 갖는 대상의 표상은 ‘특정 역할(role)에 묶인(bound) 점유자(filler)’, 다시 말해 특정 역할을 맡은 점유자로 보고 있다. 이것을 구체적으로 알아보기 위해 맥클러랜드와 루멜하트의 낱말 지각 모형(McClelland/Rumelhart, 1981)에 적용해 보자. ‘JOHN’과 같은 임의의 녀 자짜리 낱말을 표상한다고 생각해 보자. 이 때 임의의 낱말에서 각 글자가 위치를 점유할텐데 그 때 글자를 표상하는 데 쓰이는 유닛(unit)들을 점유자 유닛이라 부른다. ‘JOHN’과 같은 경우는 그 단어가 ‘J’, ‘O’, ‘H’, ‘N’에 의해 점유되어 있으므로 ‘J’, ‘O’, ‘H’, ‘N’ 각각이 점유자 유닛이 될 것이다. 그리고 글자의 위치들은 글자가 행하는 역할이라고 생각할 수 있으므로 그 위치들을 표상하는 역할 유닛을 생각해 볼 수 있을텐데 이 경우는 각 위치-첫번째, 두번째...-가 역할 유닛들이 될 것이다. 그리고 마지막으로 이 두가지 유닛들을 통해서 우리는 특정 글자를 표상할 수 있다. 곧 ‘첫번째 위치에 있는 J’ 하는 식으로 말이다. 그것을 표상하는 데 쓰이는 유닛이 묶음 유닛(binding unit)이다. 곧 이 유닛은 역할에 묶인 점유자(a filler bound to a role)를 표상한다.

이해를 돕기 위해 ‘세번째 자리에 있는 H’를 표상하는 것을 그림 1로 나타내 보자. 이 그림에서 왼쪽 가장자리에 있는 유닛이 점유자 유닛이고 맨 아래쪽의 유닛이 역할 유닛이며 나머지 유닛들은 묶음 유닛이다. 여기서 검은 동그라미는 유닛이 활성화되었음을 뜻한다.

이제 텐서 곱 개념을 끌어들일 때이다. 점유자 f 를 역할 r 에 묶는다고 하자. 이 때 역할 r 의 표상은 벡터 공간 V_R 에서의 벡터 \mathbf{r} 이다. 또 점유자 f 의 표상은 벡터 공간 V_F 에서의 벡터 \mathbf{f} 이다. 그러면 f/r 이 역할 r 에 점유자 f 를 묶음을 뜻한다면 f/r 의 표상은 텐서 곱 벡터 공간 $V_B = V_F \otimes V_R$ 에서의 텐서 곱 벡터 $\mathbf{f}/\mathbf{r} = \mathbf{f} \otimes \mathbf{r}$ 이다. 위 글자 표상 모형에서 생각해 본다면 그림 1의 왼쪽 가장자리에 있는 유닛들이 한 활성화 벡터를 이루고 아래쪽에 있는 유닛들이 다른 활성화 벡터를 이룬다고 볼 수 있는데 일단 그렇게 본다면 이 벡터들의 텐서 곱 벡

터-나머지 유닛들-가 알맞는 위치에 있는 알맞는 글자를 표상한다고 주장할 수 있다.

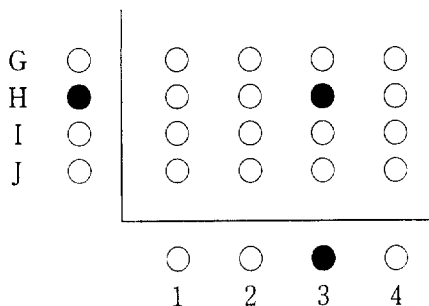


그림 1. 세번째 자리에 있는 'H'의 표상

이제 이 개념들을 이용하면 “성자는 귀남이를 사랑한다”의 구성적 구조를 표상하기 위해 먼저 ‘성’, ‘자’, ... 각 글자를 표상할 수 있다. 첫 글자 ‘성’ 만 보기로 들어보면 그림 2와 같이 될 것이다.

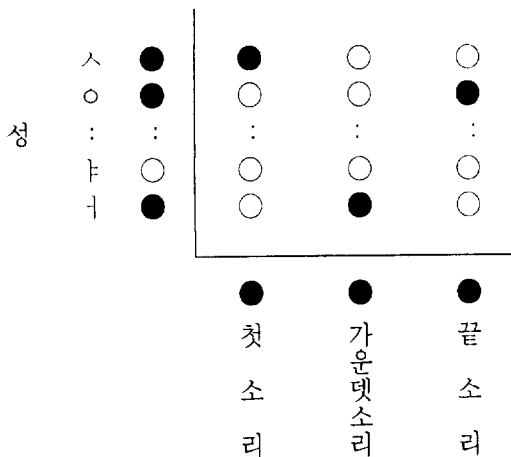


그림 2. 글자 '성'의 표상

이런 식으로 각 글자를 표상한 다음 그 글자를 통해 낱말을 표상하기 위해 우리의 시각을 더 넓혀야 한다. 곧 우리 네트워크에서는 주어진 위치에 있는 글자의 표상뿐만 아니라 낱말 자체의 분산 표상까지 원한다. 이 때 도입되는 개념이 벡터 합, 패턴 중첩(superposition)이다. 만약 두 글자가 연결주의 네트워크에서 어떤 활성화 패턴에 의해 각각 표상되었다면 이 두 글자를 접속(conjunction)한 복합 글자, 또는 낱말의 표상은 두 개별 패턴을 중첩시켜서 나온 패턴이다. 이것은 각 위치에 있는 각 글자에 해당하는 벡터들의 텐서 곱을 그냥 더하기만 하면 된다. 곧 $f_{성}$ 과 $f_{자}$ 가 각각 글자 '성'과 '자'의 점유자 벡터이고 r_1 과 r_2 가 각각 첫번째 위치, 두번째 위치를 가리키는 역할 벡터라면 '성자'의 벡터값은

$$f_{성} \otimes r_1 + f_{자} \otimes r_2$$

이다.

이렇게 하여 우리는 '성자', '귀남', '사랑한다'의 표상을 갖게 되었다. 이제 "성자는 귀남이를 사랑한다"의 표상을 이 요소 표상들로부터 만들어 낼 수 있다. '성자'와 '귀남'은 이 새로운 표상을 만들기 위해 새로운 점유자 유닛이 된다. 이런 점에서 벡터 곱, 또는 벡터 합은 되부름적 과정이다. 그러면 그 점유자 유닛과 결합하는 역할 유닛은 각각 '행위자'와 '피동자'가 될 것이다. 그래서 '행위자로서의 성자'의 표상은 $f_{성자} \otimes r_{행위자}$ 라는 텐서 곱이 될 것이고 '피동자로서의 귀남'의 표상은 $f_{귀남} \otimes r_{피동자}$ 라는 텐서 곱이 될 것이다. '행위자로서의 성자'와 '피동자로서의 귀남'을 복합시킨 표상을 만드는 것은 이제 어렵지 않다. 이 두 텐서 곱을 중첩, 벡터 합하면 될 것이다. 우리가 알고자 하는 최종적인 표상, 곧 "성자는 귀남이를 사랑한다"는 이 벡터 합에 '사랑한다'에 해당하는 벡터 값—아마도 $f_{사랑한다} \otimes r_{행위}$ —을 더하면 될 것이다. 곧 '성자는 귀남을 사랑한다'의 벡터 표상은

$$[f_{성자} \otimes r_{행위자}] + [f_{귀남} \otimes r_{피동자}] + [f_{사랑한다} \otimes r_{행위}]$$

와 같이 될 것이다(최 훈, 1994a, 85-9쪽 참조).

이상과 같은 텐서 곱 표상은 포더와 필리신의 반론을 벗어날 수 있을 것 같다. 한 구조화된 대상의 텐서 곱 표상은 그 대상의 요소들 각각에 대한 점유자/역할 묶음을 표상하는 벡터들을 합함으로써 만들어질 수 있기 때문이다.

텐서 곱 표상에 대해서 두 가지 특징을 지적해야겠다. 첫째로 텐서 곱 표상은 단순한 부분/전체 관계는 아니라는 것이다. 초기의 국부적인(local) 연결주의 모형에서처럼 특정한 유닛들이 특정한 고정된 표상 내용에 부여되기만 하고 그 외에 구조를 반영할 만한 아무런 장치를 갖추고 있지 못한 경우를 생각해 보자. 그렇다면 ‘성자’, ‘귀남’, ‘사랑한다’에 각각 해당하는 특정 마디(node)들이 있을 텐데 그 마디들이 동시에 발화한다고 할 때 ‘성자는 귀남이를 사랑한다’는 표상과 ‘귀남이는 성자를 사랑한다’는 표상을 구분할 방법이 도대체 없고 그래서 체계성이 설명 안 되는 것이다. 그 때의 전체 표상과 요소 표상은 단순히 부분/전체의 포함 관계일 뿐이고, 임자말[주어], 풀이말[술어] 등과 같은 풍부한 구문론적 요소 관계를 반영하지는 못하는 것이다.²⁰⁾ 그런데 이미 보았던 것처럼 텐서 곱 표상의 점유자/역할 묶음은 한 표상의 요소 표상을 나타내줄 뿐만 아니라 임자말, 부림말[목적어], 움직씨[동사]와 같은 구조적 관계까지도 보여줄 수 있다.

또 다른 특징은 텐서 곱 표상을 풀어 헤쳐 그 요소 표상들을 되찾을 수 있다는 것이다. 곧 요소 표상을 통해 구조화된 복합 표상의 텐서 곱 표상을 합성할 수 있을 뿐만 아니라, “성자는 귀남이를 사랑한다”를 표상하는 벡터에서 ‘성자’와 ‘귀남’, 그리고 ‘사랑하기’를 표상하는 벡터를 분명히 일반적이고 효율적이고 믿을 만한 방식을 통해서 찾을 수 있는 것이다.²¹⁾ 나는 I 절에서 구성적 구조는 ‘그 요소를 복

20) Horgan/Tiensen(1992), 206-7쪽 참조.

21) 스몰렌스키는 이것을 보장해 주는 다음과 같은 정리를 말하고 있다.

$\mathcal{P}F/R$ 를 역할 분석에 의해 생긴 텐서 곱 표상이라고 해보자. 구조 s 에서 결합된 역할을 표상하는 벡터가 선형적으로 독립적이라고 가정하자. 그러면 각 역할은 완전히 정확하게 분해될 수 있다. 곧 각 결합된 역할 r 에 대해서 s 를 표상하는 벡터 $\mathcal{P}F/R(s)$ 로부터 r_i 에 결합된 점유자 f_i 를 표상하는 벡터 f_i 를 얻는 과정이 있다.

합 표현에서 분해해 내고 다시 복합 표현으로 합성해 낼 수 있는 일 반적이고 효율적이고 믿을 만한 방식'을 만족하는 조건이어야 한다고 말했다. 텐서 곱 표상은 이 조건을 만족함을 알 수 있다.

Ⅲ. 참된 구성적 구조

3.1. 참된 구성적 구조가 되기 위한 필요충분조건

이제 가장 중요한 문제를 짚고 넘어가야 할 순서이다. 우리는 이 글에서 내내 연결주의는 강한 의미의 구성적 구조는 아니지만 약한 구성적 구조는 보임으로써 고전주의자들이 요구하는 바를 만족시킬 수 있음을 알았다. 그런데 고전주의자들이 계속 트집을 잡을 수가 있다. 구성적 구조라는 기준의 원정 경기에 참가한 것만 해도 연결주의자는 고전주의자에게 상당히 양보한 건데 고전주의자는 거기에 그치지 않는다. 그들은 구성성이 혹시 강한 의미와 약한 의미로 나누어진다고 하더라도 참된 의미의 구성성은 강한 의미의 구성성이어야 한다고 고집을 피울 수 있는 것이다. 결국 연결주의자가 말하는 약한 구성성은 그런 게 있다고 하더라도 올바른 종류의 구성성이 아니라는 것이 그들의 하고 싶어 하는 얘기이다.

이제 우리는 이 짐까지 벗어버려야 포더와 그의 동조자들의 공격을 완전히 막을 수 있다. 그러기 위해서는 약한 의미의 구성성이 참된 의미의 구성성임을 보여 주어야 한다. 참된 구성성(real compositionality)이란 무엇일까? 포더가 스몰렌스키의 요소 구조를 공격하기 위해 '참된 요소성'(real constituency)이라는 말을 쓰긴 하지만 과연 어느 정도의 강도를 갖는 구성성이 참된 구성성인지는 구

이 정리와 그에 대한 증명은 Smolensky(1990)의 186쪽을 보라. 한편 이 정리에서 벡터들이 선형적으로 독립적이지 못하다면 요소 벡터들을 되찾을 수 있는 가능성은 정도의 문제로 남는다고 Horgan과 Tiensen은 말하고 있다. 곧 선형적인 독립성이 작을수록 요소가 되는 점유자, 역할 벡터를 되찾을 수 있는 수학적 가능성은 그 정확도가 줄어든다는 것이다(Horgan/Tiensen, 1992, 208쪽).

성성이 요청되는 상황을 되새겨 보는 게 옳은 방법일 것이다. 그래서 우리에게 구성성이 있어야만 되는 상황과 배경을 만족시킬 수 있는 구성성이면 되지 그 이상의 조건도 그 이하의 조건도 우리에게서 짐이 될 뿐이다.

우리는 I 장 2절에서 구성성이 자연 언어에 요청되는 근거가 생산성 논변 또는 체계성 논변을 통해서라고 말했다. 그 때 그 논변에서 구성성을 위해서는 (가) '요소적 구조'가 있어야 하고 (나) '그 요소를 복합 표현에서 분해해 내고 다시 복합 표현으로 합성해낼 수 있는 일반적이고 효율적이고 믿을 만한 방식'이 있으면 된다고 하였다. 일반적인 경우에 구성 또는 합성에 요구되는 것은 구성 요소가 있고 구성 요소가 주어졌을 때 복합 표현을 만들어 내고 그 복합 표현을 다시 요소 표현으로 분해할 수 있는 체계적인 방법을 지닌다는 것이 전부일 것 같다. 언어의 생산성을 설명하기 위해서 구성성을 필요로 하고 구성성을 설명하기 위해서 요소성과 그 요소에 민감한 구조가 필요하다. 어떤 체계에서 일단 요소가 주어지고 그 요소에 어떤 뒤부름적 규칙을 적용해서 복합적인 표현을 만들 수 있다면 그 체계의 생산성을 설명하기 위한 최소한의 조건은 만족된 셈이다. 그 이상의 조건은 덤일 뿐이다. 그렇다면 그 요소를 글자 그대로 보존해야 한다는 조건은 생산성을 설명하기 위해서는 꼭 필요한 조건이 되지 못한다. 그 조건이 없어도 언어의 생산성을 설명할 수 있지 않는가? 추상적인 요소 관계를 만족하는 표현을 만든다는 관점에서는 표현 자체에서 요소들의 사례들을 보존해야 한다는 필연성은 없다. 다만 중요한 것은 그 표현이 구성적인 구조의 기능을 보인다는 점일 뿐이지 않겠는가? 만일 그렇다면 약한 의미의 구성적 구조도 참된 의미의 구성적 구조로 인정해야만 한다.

이렇게 대충 (가)와 (나)의 조건만으로 생산성과 체계성을 설명할 수 있다는 설명을 했지만 (가)와 (나)가 참된 구성적 구조의 필요 충분 조건이 되기에는 더 엄격한 설명이 필요할 것이다. 대부분의 필요 충분 조건의 증명이 그렇지만 한 쪽 방향의 증명은 쉽다. 이 경우에도 조건 (가)와 (나)가 참된 구성적 구조가 되기 위한 필요 조건이라는 것을 보이기는 간단한 일이다. 참된 구성적 구조가 만족되어서 표

상의 생산성과 체계성을 올바르게 설명할 수 있는 경우에는 당연히 (가)와 (나)의 조건이 만족되어 있을 것이다. 조건 (가)와 (나)는 포더와 그의 동조자들도 인정하는 최소한의 조건임을 잊지 않았을 것이다. 그런데 문제는 그 반대 쪽 방향이다. 과연 조건 (가)와 (나)만을 만족해서 참된 구성적 구조를 보인다고 할 수 있을까? 곧 표상의 생산성과 체계성을 올바르게 설명할 수 있는 충분 조건일까? 포더와 그의 동조자들이 조건 (가)와 (나) 말고 구성적 구조에 필요하다고 생각하는 조건은

(다) 한 표현이 다른 표현의 요소일 수 있는 경우는 다른 표현이 구현될 때마다 그 표현이 구현되는 경우일 때만이다.

임을 우리는 알고 있다. 그리고 이 조건이 말하려고 하는 바는 단순 기호와 복합 기호 사이에는 구체적인 의미에서 곧 글자 그대로의 의미에서 부분-전체 관계가 있고 단순 기호는 그 단순 기호가 나타나는 어떤 복합 기호에서도 글자 그대로 보존된다는 것이었음도 알고 있다. 바로 포더와 그의 동조자들은 자신들이 강조하는 고전적인 요소성을 지닌 구조여야만 올바른 구성적 구조를 구현할 수 있다고 생각하는 것이다. 따라서 그들에 따르면 강한 의미의 구성적 구조만이 참된 구성적 구조이고 표상은 (강한) 언어성을 띠어야만 하며 그러기에 약한 의미의 구성적 구조를 보이는 연결주의 모형은 구성적 구조를 제대로 보인다고 하기가 어렵다. 약한 의미의 구성적 구조는 엄격히 말해서 그들의 논의에 필요한, 곧 표상에 필요한 구성적 구조가 아니니까.

글쓴이가 조건 (가)와 (나)면 참된 구성적 구조가 되기에 충분하다는 것을 증명하기 위해서 쓰는 방법은 포더와 그의 동조자들이 충분한 조건이 되기 위해서는 더 만족해야 한다고 생각하는 조건 (다)가 사실은 불필요한 군더더기 조건이라고 주장하는 것이다. 다시 말해서 참된 구성적 구조에 충분한 조건은 앞에서 말한 조건 (가)와 (나)면 되지 (다)까지 필요하지는 않은 것이다. (더 나아가 그게 있으면 포더와 그의 동조자들의 주장에서 모순을 일으킨다.) 다시 이것을 보이

기 위해서는 그들이 참된 구성적 구조에 (다)까지 요구하는 까닭이 무엇인지 알아내서 그 근거가 희박함을 지적하는 방식을 쓸 것이다. 조건 (다)가 구성적 구조를 참되게 하기 위해서 있어야만 하는 이유가 그럴 듯하지 못하다면 우리는 그 조건까지 받아들일 필요는 없게 되고, 그렇다면 우리는 (어쩌면 우리가 아는 한) 조건 (가)와 (나)만 있으면 참된 구성적 구조가 되기에 충분한 것이다. 따라서 조건 (가)와 (나)는 참된 구성적 구조가 되기 위한 필요 충분 조건이 된다.

그럼 왜 포더와 그의 동조자들은 조건 (다), 구체적으로는 고전적인 요소성을 구성적 구조에 고집하는 것일까? 왜 요소가 어떤 풀바꿈 속에도 글자 그대로 보존되어야 한다는 조건이 구성적 구조가 되기 위한 조건에 필수적으로 들어가야 하는가? 글쓴이가 보기에는 두 가지 까닭에서 그런 것 같다. 첫째로 포더 등이 사고에 구성성이 있다고 생각한 동기는 I장 말했 듯이 자연 언어가 구성성을 가짐에 준해서 생각했기 때문인데 우리 자연 언어에서 흔히 보이는 요소 표현과 복합 표현의 관계는 조건 (다)를 만족하는, 곧 요소 표현이 분해와 합성의 과정에서 글자 그대로 보존되는 부분-전체 관계이기 때문이다. 둘째는 구체적으로 사례화된, 다시 말해서 구현된 요소 표현만이 인과적인 힘을 가지고 있다고 생각하기 때문이다. 그렇지 못한 요소는 가상의 것이고 아무런 효과를 가져오지 못한다. 이 글에서는 둘째 논변에 대해 답변을 하고 그 이유가 근거 없음을 보임으로써 (다)의 조건이 굳더더기 조건이라고 주장하겠다.²²⁾ 그 다음에 3.5절에서는 더 나아가 그 조건이 있으면 자신들의 주장과 모순을 불러 일으킴을 지적하겠다.

3.2. 포더와 맥래플린의 의심

나는 2.2절에서 텐서 곱 표상을 통해 연결주의의 표상도 요소 구조를 가질 수 있고 그러기에 구성적 구조를 보일 수 있다고 주장했다. 그런데 포더와 그의 동조자들은 연결주의에서 말하는 요소는 참다운 요소가 아니라고 주장한다. 더 정확히 말하면 연결주의 모형의 활성

22) 첫째 논변에 대해서는 최 훈(1994a), 3.3절 참조.

화 벡터는 고전적인 요소(classical constituency)를 갖지 않는다. 포더와 맥래플린은 고전적 요소성에 대해서 이렇게 설명한다.

고전적인 견해는 정신적인 표상의 구문론이 자연 언어 문장의 구문론과 다음과 같은 점에서 같다고 주장한다: 둘 다 우리가 고전적인 요소라고 부를 것으로부터 만들어진 복합적인 기호들을 포함한다. ..우리는 이렇게 가정한다. 한 쌍의 표현 E1과 E2에 대해서 E1이 E2의 고전적 요소인 것은 E2가 사례화할(tokened) 때마다 E1이 사례화할 때 뿐이다(Fodor/Pylyshyn, 1988, 333쪽).

‘고전적’이라는 형용사가 붙은 요소와 그렇지 않은 요소가 갖는 차이점은 우리가 이미 고전적 요소의 정의를 통해 알고 있듯이 그 요소가 자신이 나타나는 표현 안에서 구체적으로 사례화하느냐 그렇지 않느냐의 차이임을 알 수 있다. 이 차이가 어떤 결과를 가져오기에 포더와 그의 동조자들은 참다운 요소가 되느냐 못되느냐의 결과까지 끌어 낼 수 있다고 생각하는가? 그것은 정신 과정의 관점에서 복합적인 기호의 고전적인 요소들은 사실상 복합적인 기호가 사례화하는 데 인과적으로 영향을 끼치는데 비해 연결주의에서 말하는 텐서 곱과 중첩 벡터와 같은 요소들은 인과적인 상태 자체를 가질 수 없다는 점 때문이다. 한 마디로 말해서 구체적으로 일어난 일은 그것이 일어나 있다는 자체만으로도 무슨 일인가를 일어나게 하는 데 참여하고 있지만 단지 가상의 것은 무슨 일이든 일어나게 할 수는 없다. 포더와 맥래플린에 따르면,

...복합 기호의 고전적인 요소들은 복합 기호가 사례화될 때 인과적인 영향을 결정하는 데 이바지한다. ..그러나 거칠게 말해서 육이 “ 3×2 ”로 표상될 수 있다는 사실은 그것만 가지고는 육이 “6”으로 표상되는 셈틀[컴퓨터](또는 뇌)에서 일어나는 인과적 과정에 영향을 끼칠 수 없다. 단순한 반사실적 표상들은 인과적인 영향력을 갖지 못한다. 실제로 사례화되는 표상들만이 인과적인 영향력을 가질 뿐이다(Fodor/ McLaughlin, 1991, 346쪽; 강조는 글쓴이가).

정신적인 표상이 인과적인 힘을 갖기 위해서는 그 기호들이 사례화되어야 한다는 것이다.

포더와 맥래플린의 주장을 정리해 보면 다음과 같다.

1. 올바른 구성적 구조는 복합 표현에서 요소가 되는 표현이 구체적으로 사례화되어야 하고 어떤 복합 표현에서도 그 사례가 보존되어야 한다.
2. 그래야만 하는 이유는 실제로 사례화하는 요소만이 인과적인 영향력을 행사할 수 있기 때문이다.
3. 단지 가상의 것은 인과적인 힘을 가질 수 없다. 그런데 연결주의에서 말하는 요소는 가상의 것이다.
4. 그런데 실제로 사례화하는 기호적인 구조를 우리 마음도 지녀야 한다.
5. 그러므로 사고 언어의 구성적 구조는 구체적으로 사례화되는 고전적인 요소를 포함하고 있어야 하고 그런 특징을 보여 주지 못하는 연결주의 모형은 올바른 모형이 아니다.

이와 같은 지적은 텐서 곱 표상에 대해서 틀린 말은 아니다. 포더 등은 텐서 곱 표상에서 나오는 점유자와 역할 유닛의 활성화 상태가 가상적(imaginary)이라고 주장한 것이다(Fodor/McLaughlin, 1991, 344쪽). 이게 무슨 말인가 들어 보기 위해 ‘성자’를 표상하는 기계를 다시 생각해 보자. 낱말 ‘성자’를 표상하는 중첩 패턴은 기계에서 실제로 실현되는 활성화 벡터일 것이다. 그러나 글자 ‘성’을 표상하는 활성화 벡터는 단지 가상적이다. 또 ‘첫번째 위치에 있는 글자임’을 표상하는 활성화 벡터도 가상적이다. 그리고 또 ‘첫번째 위치에 있는 글자 “성”’을 표상하는 텐서 곱 벡터도 마찬가지로 가상적이다. 가상적이지 않은 활성화 패턴은 ‘성자’를 표상하는 중첩 벡터일 뿐이다. 포더와 맥래플린이 이렇게 주장하는 근거는 딱 하나뿐이다. ‘성자’를 표상하는 벡터 표상에서 실제로 사례화된 것은 ‘성자’를 표상하는 중첩 벡터 밖에 없기 때문이다. 그리고 글자 ‘성’이 가상의 것인 이유는 그것이 텐서 곱과 패턴 중첩을 통해서 낱말 ‘성자’를 표상하는 중첩 벡터를 이루게 하지만 실제로 사례화되지는 못했기 때문이다.

3.3. 의심풀기

포더와 맥래플린이 텐서 곱 표상의 요소에 대해 품은 이러한 의심을 어떻게 풀어주어야 할까? 나는 두 가지 방향에서 이 의심을 풀어보려고 한다. 첫번째는 그들이 어떻게 해서 그런 의심을 가지게 되었는가를 뒤쫓아가 보아, 그 의심의 근거가 회박함을 지적할 것이다. 두번째로는, 앞 절의 주장 2와 3에 주목해 본다. 과연 실제로 사례화하는 요소만이 인과적인 영향력을 행사할 수 있고, 단지 가상의 것은 인과적인 힘을 가질 수 없는가? 텐서 곱 모형의 경우를 다시 생각해 보며 그에 대한 반례를 제시해 보자.²³⁾

첫번째로, 왜 포더와 맥래플린은 실제로 사례화되지 못한 요소는 가상적이라고 생각할까? 포더와 맥래플린은 이렇게 말하고 있다.

텐서 곱과 중첩 벡터의 요소들은 고전적 요소와 다음과 같은 점에서 차이가 난다: 복잡한 고전적 기호가 개체화될 때 그것의 요소들도 개체화된다. [이에 견주어] 텐서 곱 벡터 또는 중첩 벡터가 개체화될 때 그것의 요소들은 (우연한 경우는 제외하고는) 개체화되지 않는다. 이 차이점이 함축하는 바는 복합적 기호의 고전적 요소들은 그것이 개체화하는 데 인과적인 영향력을 행사하지만...텐서 곱과 중첩 벡터의 요소들은 그런 인과적인 상태를 갖지 못한다는 점이다(같은 글, 345쪽).

여기서 고전적인 요소란 복합 표현이 사례화(개체화)될 때 동시에 사례화되는 요소를 말한다(같은 글, 333쪽). 이를테면 ‘성자는 귀남을 사랑한다’가 사례화될 때마다 ‘성자’가 사례화되므로 뒤의 것이 앞의 것의 고전적 요소이다. 그렇다면 위 인용문에서 텐서 곱 표상에 나오는 요소와 고전적인 요소의 차이점은 사례화되었느냐 그렇지 못하느냐의 차이뿐이다. 다시 말해서 구현되었느냐 그렇지 못하느냐의 차이

23) 나는 최 훈(1994a)에서 이외에도 포더와 그의 동조자들이 우리 정신적 표상에 구체적으로 사례화하는 기호를 도입하게 되는 근거로서 그 기호가 있음으로 해서 우리 사고에 있다고 가정되는 인과적 측면과 의미론적 측면 사이의 연관성을 잘 드러낼 수 있다는 점을 생각하고 있다는 것을 지적했다(117-122쪽). 지금 논의를 집중시키기 위해 이 점은 여기서 생략하자.

뿐이다. 이 차이점이 함축하는 바는 정신 과정에서 고전적인 요소는 자신이 나타나는 복합 표현이 사례화될 때 동시에 사례화됨으로써 복합 표현이 사례화되는 데 인과적인 영향을 끼치지만 텐서 곱의 표상들은 전혀 그런 인과적인 힘을 쓰지 못한다는 것이다. 자신이 속한 더 큰 표현에서 공간적인 위치를 한 자리 차지하고 있다는 게 그들이 말하는 인과적인 유효성의 전부인 것 같다.

이제 우리는 포더/맥래플린과 다음과 같은 문답을 ‘가상적으로’ 주고 받는다.

물음: 왜 활성화 상태가 가상적이 아닌 실제적일 것을 요구하는가?

대답: 실제적인 활성화 상태여야 인과적인 힘을 갖기 때문이다. 그렇다.

물음: 그럼 왜 인과적인 힘이 있어야 한다고 생각하는가?

대답: 우리는 “체계성 문제에 대한 고전적인 해결은 정신적인 표상들의 요소들이 인과적인 역할들을 갖는다고 전제한다. 다시 말해서 그것들은 정신 과정(mental process)에서 세력을 끼치는 범위를 제공한다”(같은 글, 346쪽)고 말했다. 간단하게 말하자면 인과적인 힘이 있어야 되는 까닭은 그래야만 정신 과정에 세력을 끼치기 때문이다.

물음: 인과적인 힘이 있어야 된다는 것이나 정신 과정에 세력을 끼쳐야 한다는 것이나 결국은 같은 말을 되풀이한 것이 아닌가? 그럼 다시 묻는다. 왜 정신 과정에 세력을 끼쳐야 하는가?

대답: 우리는 방금 한 말에 바로 이어서 “따라서 기호의 고전적인 요소들은 복합적인 기호가 사례화할 때 인과적인 영향을 결정하는 데 이바지한다”고 말했다. 곧 정신 과정에 세력을 끼쳐야 하는 까닭은, 다시 말해서 인과적인 영향을 행사해야 하는 까닭은 사례화하기 위해서이다.

결국 우리가 알 수 있는 것은 이것 뿐이다. 곧 실제로 사례화해야만 인과적이 되고 인과적이 되면 실제로 사례화한다. 실제로 사례화하는 표상들만이 인과적인 영향력을 가질 뿐이다.

요소 표현이 복합 표현에 인과적인 영향을 끼쳐야 함은 옳은 지적

이다. 그렇지 않고서는 복합 표현과 요소 표현은 아무 관계도 없고 그 사이에는 포더와 필리신의 말을 빌리자면 “이름표를 써서 속인 허구”(Fodor/Pylyshyn, 17쪽)일 뿐이기 때문이다. 그러나 그 인과적인 힘이라는 게 어떤 형식으로 나타나야 하는가? 포더가 이에 대해서 위 인용문에서처럼 그 인과적인 힘이라는 것은 구체적인 사례화에서만 생기는 것이라고 대답한다면 그건 고전적인 요소성이 필요한 이유에 대해 아무런 새로운 대답을 하지 못한 것이 되고 만다. 왜냐하면 고전적인 요소성이 필요한 이유를 인과적인 영향력으로 설명했고 다시 인과적인 영향력을 구체적인 사례화에만 국한시킨다면 다시 고전적인 요소성으로 돌아가는 악순환을 범할 뿐이기 때문이다.

그런 어리석음을 범하지 않으려면 포더와 그의 동조자들은 부분과 전체 사이에 구체적인 사례화 외에 다른 인과적인 관계가 있어도 된다고 허락해야만 할 것이다. 그렇지 않는다면 사례화해야만 인과적인 힘을 가질 수 있다는 주장에 대해서 새로운 근거를 대지 못하고 인과적인 힘은 사례화하는 경우에만 쓸 수 있다는 주장에서만 계속 맴도는 셈이 된다. 그렇다면 우리는 포더와 그의 동조자들이 꼭 사례화해야만 인과적인 힘을 발휘할 수 있다는 데에 대해 설득력있는 설명을 하지 않는 이상, 사례화하느냐 안하느냐에 주목할 것이 아니라 요소가 되는 표현 또는 벡터가 인과적인 역할을 하느냐 하지 못하느냐 하는 문제에 관심을 가져야 한다. 그래서 우리는 다른 인과적인 관계의 가능성을 생각할 수 있다. 포더와 맥래플린의 말마따나 정신적인 표상의 요소들이 정신 과정에 세력을 끼치는 범위를 제공하려면 정신 과정에서 그 요소들이 꼭 사례화하는 방법을 통하지 않더라도 전체에 어떤 형태로든 영향을 끼치면 될 것이다. 요소들 없이는 전체가 있을 수 없다는 것을 보여 주면 요소는 전체에 대해 영향력을 갖는다. 그런 의미에서 전체가 부분의 함수가 되는 경우도 부분이 전체에 영향을 끼치는 경우이다. 전체는 부분의 값이 변화함에 따라 그 값이 변화하고 이걸 바로 부분이 전체에 영향을 끼친다는 점이 아니고 무엇이겠는가?

이렇게 보듯이 연결주의 모형의 요소도 인과적 영향력을 행사할 수 있다. 그러나 그 요소는 고전주의에서 말하는 고전적 요소성과 같

은 유형의 인과적 유효성은 아니다. 그리고 우리는 그런 고전적 요소성에 늘어붙어 인과적 유효성을 찾을 까닭이 없음을 이미 살펴보았다. (고전주의와 다른 방식으로 인과적 유효성을 지닌다는 점이 다음 절의 구현 논의에서 중요한 구실을 할 것이다.)

이제 나는 고전적 요소성과 같은 방식은 아니지만 전체 값이 부분 값의 함수라는 의미에서 연결주의의 요소가 인과적 영향력을 행사할 수 있음을 텐서 곱 표상에서 살펴보아 포더와 맥래플린의 의심을 푸는 두번째 작업을 하려고 한다.

텐서 곱 표상과 같은 연결주의에서 표상은 분석나무(parse tree)와 같은 이산적인 구조로부터 벡터 공간으로의 짝짓기(mapping)라고 하는데 결국 그 표상은 어떤 고차원 공간의 점에 대응한다고 볼 수 있다(Smolensky, 1991; Churchland, 1989 참조). 그런데 그 점이 공간에서 실제로 어디에 위치해 있느냐는 그 장소에서 벡터가 갖는 특정 값에 의해 결정된다. 다시 말해서 특정 내부 구조에 의해 결정된다. 표상을 일단 공간 속의 점으로 생각한다면 그 점들이 서로 어떤 관계에 있느냐, 다시 말해서 그 점들 사이의 '거리'가 얼마인가를 생각하는 것은 자연스러운 일이다. 이제 우리는 표상들 사이의 체계적인 관계가 전체는 부분의 함수라는, 곧 전체는 부분에 의해 결정된다는 사실을 눈치챌 수 있다. 그 표상의 내부 구조, 곧 다른 표상과 관련해서 어디에 위치해 있느냐는 다음 두 가지 점에 의해 결정된다. 첫째 그 요소의 본성, 쉽게 말하면 벡터 공간에서의 위치. 둘째 합성 과정의 본성. 결과적으로 표상이 텐서 곱 네트워크 안에서 작동하는 방식은 기능적인 요소 관계에 의존한다. 표상의 내부 구조, 곧 벡터 공간에서의 위치는 그 요소 관계의 직접적인 함수이므로 부분이 있기에 전체가 있게 되며 따라서 텐서 곱 표상은 부분과 전체 사이에 인과적인 관계가 있음을 알 수 있다. 그렇다면 고전적인 요소성에서처럼 구체적인 사례화만 인과적인 관계의 조건으로 인정해서는 안되고 이는 다시 구체적인 사례화가 구성적 구조의 조건에서 꼭 필요한 조건의 항목이 될 수 없음을 말해 준다.

스몰렌스키가 든 원자 상태의 표상이 이 점을 이해하는 데 도움이 될 것이다. 원자의 상태는 추상적인 벡터 공간에서의 벡터에 의해 표

상된다고 한다. 이 벡터는 다음과 같이 결정된다. 곧 전자의 상태(‘회전’)는 벡터(V_s)에 의해 표상되고, 원자 속의 전자의 궤도는 다른 벡터에 의해 표상된다. 그리고 원자 속에 위치하는 전자는 이 두 벡터들의 텐서 곱($V_s \otimes V_o$)에 의해 표상된다. “원자 전체는 각자 자신의 궤도에 위치해 있는 특정 전자를 표상하는 벡터들의 합 또는 중첩인 벡터에 의해 표상된다”(Smolensky, 1991, 227쪽, 주 13). 이 경우 원자 벡터값이 각자 자신의 궤도에 위치해 있는 각 전자 벡터값들의 합의 함수가 아니라고 말할 수 없을 것이다.

이 절의 결론은 이렇다. 한 마디로 말해서 텐서 곱 표상의 요소들은 인과적으로 무능하지 않다.

3.5. 새로운 딜레마

포더와 그의 동조자들은 이제 이렇게 대꾸할 수 있다. “그래, 강한 의미의 구성적 구조가 우리 언어에서 단 하나뿐인 구성적 구조는 아니다. 그러나 우리의 인지 표상은 분명히 강한 의미의 구성적 구조를 따른다.” 이렇게 포더와 그의 동조자들이 강한 의미의 구성성만을 계속 주장하고 단순히 기능적인 구성성을 인정하지 않는다면 다음과 같은 논변으로 궁지에 몰 수 밖에 없다. 곧 앞절까지는 조건 (다)가 굳 더더기임을 주장했다면 여기서는 조건 (다)가 적어도 그들의 체계 안에서는 모순을 불러 일으키므로 그들이 일관되게 주장해서는 안되는 것이라고 말하련다.

포더와 그의 동조자들은 연결주의를 딜레마에 빠뜨리려고 했다. 이제 그들도 단순히 기능적인 약한 의미의 구성성을 인정하느냐 안하느냐 하는 문제에서 딜레마에 빠질 수 있다. 만일 그들이 단순히 약한 의미의 구성성을 인정하게 되면 이 글이 본디 노리던 바대로 바로 연결주의 체계의 표상은 구성적인 구조를 갖게 된다. 그래서 그들의 연결주의 공격은 무위로 끝나고 만다. 그렇다면 그들이 단순히 약한 의미의 구성성을 인정하지 않게 된다면 그리고 강한 의미의 구성성만 인정한다면 어떻게 될까? 내 생각에는 그들이 스스로 나눈 인지/구현 차원 구분에 어긋나는 모순을 저지르게 된다. 무슨 말이고 하니 포더

와 그의 동조자들이 마아(Marr)를 따라 채택하는 차원 구분에 의하면(‘글을 맺으면서: 새로운 인지 모형?’ 참조) 정신 표상이 구성적으로 구조화되었다고 주장할 때 그것이 뜻하는 바는 표상은 올바른 종류의 기능적 속성을 가져야 한다는 것이다. 곧 모든 인과적 상호작용에서 그런 구조를 가진 듯이 행동해야 하고 그 행동이 실제로 하드웨어 차원에서 어떻게 이루어지는냐는 단지 구현 차원의 문제일 뿐이다. 그런데 구성성의 문제에 한해서만은 기능적인 차원을 인정하지 않고 그 기능이 글자로 구현되는 차원까지 요구한다는 것은 차원 구분을 아전 인수격으로 이용하고 있다는 비난을 면키 어렵다. 구성성의 문제도 어디까지나 표상에 관한 것이므로 그들이 정신 표상이 가져야 한다고 생각한 위 차원 구분이 구성성의 경우에도 적용되어야 하지 않겠는가?

다음을 보자:

- ① 우리의 사고와 행동은 구성적인 구조를 가지고 있다.
- ② 그 기저의 물리적 메카니즘도 구성적인 구조를 가지고 있다.

포더와 그의 동조자들은 처음에는 ①을 주장하다가 연결주의도 단순히 기능적인 수준이나마 ①을 주장할 수 있게 되자 ②까지 요구한다. 포더와 그의 동조자들은 분명히 구성적인 구조가 글자 그대로 적용되는 주장이라고 생각해서 기호 체계가 뇌에서 물리적으로 실현되어야 한다고 주장한다(Fodor/Pylyshyn, 1988, 13쪽). 이 때는 인지/구현의 차원을 구분하지 않고 있다. 그러나 우리는 포더와 그의 동조자들에게 인지/구현 차원 구분이 반연결주의 논변에서 중요한 구실을 했음을 알고 있다. 이것만 해도 포더와 그의 동조자들은 자기 모순을 범하고 있다. 그런데 설상가상으로 포더와 그의 동조자들도 인정하듯이 ②는 경험적인 사실이다. 이 글이 보여주려고 했던 바에 따르면 연결주의 모형들은 ②에 반대되는 연구 결과들을 보여줄 수 있는 가능성이 있다. 만일 그 때 가면 포더와 그의 동조자들은 또 인지/구현 차원의 구분을 들먹일 것인가?

그러나 포더와 그의 동조자들이 세워놓은 체계가 이런 비판으로

호락호락 넘어갈 것 같지는 않다. 그들 처지에서는 자신들이 고전적 요소가 글자 그대로 나타나고 본디 꼴대로 보존된다고 말할 때는 그 요소가 사례화되는 경우인데 그 사례화를 구현과 똑같은 것으로 놓고 비판을 하면 되겠는가 하는 대꾸를 할 수 있을 것이다. 분명 구현은 위에서 말한 것처럼 물리적인 차원에서 말하는 것이다. 그러나 사례화는 구현과 달리 추상적인 차원에서도 이야기할 수 있다. 곧 '귀남'이라는 낱말의 사례화는 글자로 또는 목소리로 되었는데 하는 물리적인 구현의 문제는 다루지 않으면서 추상적인 수준에서 이야기할 수 있는 것이다.

이런 대꾸는 얼른 듣기에 그럴 듯한 면을 지니고 있다. 고전적인 요소는 꼭 물리적인 구현의 차원에서 뿐만 아니라 추상적인 인지 차원에서도 말할 수 있는 것이기 때문이다. 그러나 그렇게 생각하는 것은 '구현'과 '사례화'라는 낱말의 차이에 너무 집착하는 것이다. 사례화의 경우에는 인지와 구현의 차원이 구분이 되고 그래서 인지의 차원에서도 사례화를 말할 수 있는 것은 사실이다. 그러나 우리가 이 논문 내내 살펴보았듯이 포더와 그의 동조자들의 주장이 살아남기 위해서는 그 두 차원이 동형적이어야 한다. 다시 말해서 인지의 차원에서도 구현의 차원에서와 마찬가지로 각 요소는 어떤 변형에서도 본디의 꼴을 그대로 간직하고 있어야 한다. 만약 그렇다면 인지의 차원이 구현의 차원과 독립적으로 설정되는 이유인 기능적인 다양성을 전혀 보여주지 못하고 있다. 인지 차원을 굳이 따로 분리하는 이유는 한 가지 물리적인 구현의 경우에도 다양한 알고리즘의 기능성을 보여주기 때문인데 그렇지 못하고 항상 구현 차원을 따르는 인지 차원이라면 설령 구분이 된다고 하더라도 그 구분이 적어도 우리의 논의와 관련해서는 큰 의미를 가지고 있지 못하는 것이다. 그렇다면 지금으로서는 '사례화'와 '구현'을 같게 보아 포더와 그의 동조자들을 비판하는 것이 있을 수 있는 일이 된다.

포더와 그의 동조자들이 빠진 딜레마는 정리해 보면 이렇다. 그들이 단순히 기능적인 약한 의미의 구성적 구조를 인정한다면 연결주의에 대한 공격은 옳게 되고, 인정하지 않는다면 자신들이 쳐놓은 울가미에 걸려들게 된다.

이상의 논의가 성공적이라면 약한 의미의 구성적 구조가 참된 구성적 구조가 되는 데 충분하고도 필요한 조건을 제시해 줄 수 있다. 그리고 연결주의 네트워크가 그런 구성적 구조를 보여줄 수 있다고 할 때 연결주의는 포더와 그이의 동조자들이 퍼붓는 공격의 한 짐을 벗어버릴 수 있게 된다.

글을 뺏으면서: 새로운 인지 모형?

눈부신 빛 속에서 눈을
바로 뜨기에는 시간이 좀
걸린다. 그렇다고 해서 처음
부터 철학적인 장막을 치는
것은 현명하지 못하다.

클락 (1991, 218쪽)

이 논문의 목적은 포더와 필리신이 연결주의자를 옹호한 딜레마의 한쪽 뿔에서 연결주의자를 풀어주려는 것이었다. 그 딜레마를 다시 되새겨 보자. 인지 모형이 모름지기 의미있는 인지 모형이 되려면 우리 인지가 꼭 갖는다고 생각되는 특징인 구성적 구조를 보여줄 수 있어야 한다. 그러나 연결주의 모형은 그런 특징을 보여주지 못해 적절한 인지 모형이 되기에는 너무 약한 모형이다. 설사 그런 특징을 보여준다고 하더라도 단지 고전주의 체계의 한갓 구현에 불과하여 인지 에 대한 새로운 개념이 되지 못한다. 이런 딜레마 중 우리가 이 논문에서 주로 다룬 부분은 첫번째 뿔이었다. 연결주의 모형은 인지의 구성적인 구조를 보여주지 못한다는 비판에 대해 연결주의자들은 어떤 식으로 응답을 해야 연결주의라는 아직은 새로운 모형이 의미있는 인지 모형이 되기 위한 걸음을 내딛어 싹이 제대로 틀 수 있는지 그 길을 더듬어 찾아 보았다. 거기서 밝은 대로를 발견했다면 우리는 적어도 딜레마의 한쪽 뿔에서나마 연결주의자를 해방시킬 수 있는 것이

다. 그렇다면 우리가 애초에 내세웠던 목적은 이론 셈이 된다.

그러나 여전히 딜레마의 두번째 뿔이 남아 있다. 연결주의가 설사 구성적인 구조를 보이는 모형이 된다고 하더라도 연결주의는 참된 인지 과정에 대한 설명이 아니라 구현 차원에 대한 설명일 뿐이므로 인지 이론과 상관없는 차원의 이론일 뿐이라는 것이다. 비록 이 논문의 관심사를 구성적 구조에 관한 논의에 제한했지만 그 관심이 연결주의에게 의미있는 인지 모형의 지위를 돌려주기 위한 데서 나왔다는 점을 생각해 본다면 딜레마의 두번째 뿔마저 완전히 떨쳐버려야 제대로 된 인지 모형의 입지를 지닐 수 있게 될 것이다. 그러나 이 글을 맺는 마당에서는 두번째 뿔을 피하기 위해서는 어떤 난관들을 헤쳐나가야 하는지를 귀뜸하는 정도로 그쳐야 할 것이다.

포더와 필리신은 인지 현상에 대한 의미있는 이론은 일정한 기술의 차원에서 성립하는데 그 차원은 알고리즘과 표상의 차원이라고 한다. 그런데 연결주의가 고전주의(규칙에 따른 기호 조작 모형)보다 우월하다고 알려진 점들(최 훈(1994a), 부록 4참조)은 모두 구현 그리고 이론가들이 특정 경우에 쓰곤 하는 알고리즘의 특정 실현(realization)과 관련된 특성들일 뿐이고 그러기에 인지 이론과는 상관없는 차원에서 말해지는 이론일 뿐이라고 주장한다.

포더와 필리신의 말이 옳다면 연결주의 모형은 결코 인지과학에 대한 혁명적인 새로운 토대를 제공한 것이 되지 못한다. 그것은 기껏해야 인지체계의 구현 이론일 뿐이다. 다시 말하면 일정한 정보 처리 체계의 추상적 구조가 아니라 그 물질적 기반을 다루는 이론일 뿐이다. 인지론자들에 따르면 우리가 심리학이라 할 만 한 것은 일정한 수준의 기술의 차원(level of description)에서 성립할 수 있는 것인데 그러한 기술의 차원에서만 우리가 관심을 가질 수 있는 행위에 대한 설명과 예측이 얻어질 수 있다. 그런데 포더와 필리신은 그런 기술의 차원은 구현의 차원이 아니라 알고리즘과 표상의 차원이고 연결주의는 구현의 차원에 있는 이론이라고 보는 것이다. 그렇다면 연결주의 모형은 인지 과정의 본성에 대해 원리적으로 중립적이다. 그 모형은 기껏해야 뇌(또는 아마도 이상화된 뇌와 같은 연결망)가 전통적인 인지 과학이 가설화한 과정의 유형들을 현실화한 방법을 설명하려

고 하여 고전적인 정보처리 심리학의 목표를 진보시켰다는 평가를 받을 수 있을 뿐이다. 유기체의 비표상적 상태들이 연결주의 네트워크를 구성한다고 말하는 것은 아무런 쓸모가 없다. 왜냐하면 그렇게 말하는 것은 마음이 심리적 차원, 곧 표상 차원에서 어떤 연결망이냐고 묻는 질문에 아무 것도 대답하는 것이 없기 때문이다. 연결주의 모형이 심리적 차원의 설명에 아무런 도움을 줄 수 없는 것은 분자 모형이나 양자 역학 모형, 생화학 모형이 그럴 수 없다는 것과 마찬가지로이다.

이러한 내용을 지닌 포더와 필리신의 두번째 빨에서 연결주의자를 구해줄 수 있는 길은 무엇일까? 얼른 생각할 수 있는 것은 연결주의가 과연 표상 차원의 이론이 되지 못하고 구현 차원에만 머무르는 이론일까 하는 의심이다. 우리는 머리말에서 연결주의도 표상에 개입하는 이론임을 분명히 하였다. 더욱이 이 논문이 진행되면서 그들이 미시 특성 등을 통해 표상의 구실을 모형화하려는 노력을 기울이고 있음을 보았다. 포더와 그의 동조자들이 이 부분을 문제삼지 않고 연결주의가 표상에 대해 갖는 관심을 인정한다면 연결주의에 대한 평가는 새로운 전기를 마련할 것이다. 그렇다면 그들처럼 이제 막 자라나는 연결주의를 처음부터 표상과 관계없는 차원의 이론이라고 못박는 것은 피어오르는 싹을 밟는 듯해서 올바른 태도는 아닌 것 같다. 연결주의자들은 분명히 표상에 대해 관심을 가지고 이론을 구성해 가고 있으며 더구나 그 표상은 고전주의자들이 내세우는 표상과는 차별화되는 특징들을 지니고 있어 고전주의 모형의 좋은 맞수가 되고 있다. 백보 양보하여 연결주의가 구현 차원의 이론이라고 하자. 그렇더라도 차원 구분이 그렇게 뚜렷하지 않아 그 구현 차원이란 것이 표상 차원과 겹치는 부분이 많을 수도 있다. 또 뚜렷하다 하더라도 구현 차원의 발전이 표상 차원에 영향을 주어 고전주의와 차별화되는 길을 열어주지 않는다고 장담할 수는 없다.

그러나 이런 길이 그리 평탄하지만은 않다. 연결주의자가 말하는 차원이 비록 구현의 차원에서 표상의 차원으로 올라와 고전주의자가 논의되는 차원과 같은 곳으로 들어섰다 하더라도 여전히 문제는 남는다. 고전주의자는 같은 차원이라도 그것이 나타나는 방식이 여러가지

로 있을 수 있으며 자신들이 말하는 표상은 가장 높은 수준의 차원이 라고 주장할 수도 있는 것이다. 이런 주장에 얼른 답변하지 못하는 것에 대해서는 애초에 이 논문이 거기까지 목표로 삼지 않았다는 핑계를 댈 수 있을 것이다. 이 점은 이 논문이 갖는 한계이다. 동시에 앞으로 이 논문의 연구가 계속 진행되어야 한다는 과제이기도 하다. 다만 연결주의 모형이 고전주의자가 주장하는 구현 차원에서 고전주의자들의 논의가 진행되는 표상 차원으로 올라섰다는 데서 위안을 삼아야 할 것이다. 같은 표상 차원에서도 여전히 차이가 진다고 하더라도 앞으로의 연구 결과에 따라 그 차이는 점점 더 줄어들 것이다.

그렇다면 연결주의와 고전주의를 당장 어떤 관계로 봐야 하는가에 대해 간단히 말하고 이 글을 맺어야겠다. 연결주의가 말하는 표상은 외적인 기호 체계에서 볼 수 있는 고정적이고 문맥에 독립적인 표상은 아니었다. 오히려 외적인 기호보다 훨씬 더 잘게 나누어졌고 문맥에 민감한 특징을 지닌 표상이었다. 그러나 우리에게서 어떻게 생각하면 고전주의자들이 말하는 고정적이고 문맥 독립적인 표상과 연결주의자가 말하는 유연하고 문맥에 민감한 표상을 모두 지니고 있는 것 같다는 생각이 든다. 스몰렌스키(1987)가 말한 것처럼 우리는 인지 생활에서 다음과 같은 “인지의 패러독스”에 빠져 있는지도 모른다. 그게 무엇이고 하니, 인지는 한편으로는 견고하다(hard). 논리학이나 언어의 규칙이 규정하고 있고 상황의 변화에도 고정되어 있는 기호를 우리는 필요로 하고 있다. 그러나 인지는 다른 한편으로는 유연하다(soft). 우리의 인지 능력은 많은 변화와 복잡함을 필요로 한다.

이 두 가지 선택지에서 우리가 취할 수 있는 태도는 다음과 같은 것들이 있을 수 있다. 먼저 어느 한쪽을 포기하는 것이다. 그러나 인지 현상에 분명히 드러나는 듯한 것을 버리기란 쉽지가 않다. 그렇다면 우리 뇌가 두 부분을 담당하는 기능을 두 군데로 나누어 반반씩 가지고 있다고 가정하는 것이다. 그러나 이는 신경 생리학적으로 그럴 듯 하지도 않고 또 그렇게 경계를 확실히 그을 수 있다고 가정하는 것 자체가 견고함 쪽에 우위를 두는 태도이다. 그렇다면 남은 길은 한 가지 뿐이다. 두 가지를 동시에 받아들이면서도 어느 한쪽에

우위를 두는 태도이다. 그런데 이 논문을 지금까지 읽어온 이들은 그 우위를 유연함 쪽에 두어야 한다는 것을 어렵지 않게 알아낼 수 있을 것이다.

그렇다면 우리는 왜 견고함, 곧 규칙과 기호를 쉽게 포기할 수 없는 것일까? 그에 대한 대답으로서는 뉴우턴 역학과 양자 역학의 관계를 생각해 보는 것이 도움이 될 것이다. 뉴우턴 역학과 양자 역학은 세계에 있는 대상의 행동에 대해 정확히 같은 예측을 한다. 더구나 뉴우턴 이론이 때때로 계산하기에 더 간단하다. 그 이론은 대상을 우리가 보는 방식대로 온전한 형태로 논의하고 내부 구조의 많은 부분에 대해서는 무시하기 때문에 그렇다. 그러나 많은 경우에 뉴우턴 역학은 실패한다. 우리는 그 경우에 미시 구조적인 양자 이론의 설명에 기댈 수 밖에 없다. 따라서 양자 역학은 뉴우턴 역학의 라이벌이 될 수 있고 더구나 그 경쟁에서 승리한 이론이다(Smolensky, 1988, 172쪽; Rumelhart/McClelland, 1986, 124-5쪽).

이 비유를 통해서 우리는 다음과 같은 사실을 알 수 있다. 뉴우턴 역학과 양자 역학 사이의 관계를 철저히 이해하면 우리는 거시적인 단계의 기술은 더 미시적인 단계의 기술과 견주어 볼 때 단지 대략적인 기술일 뿐이라는 것을 이해할 수 있다. 더구나 우리는 물리학에서 언제 거시 이론이 실패하고 언제 미시 이론을 필요로 하는지 안다. 이렇게 봤을 때 거시 이론은 그 이론이 미시 이론과 갖는 관계 덕분에 쓸만한 형식적 도구 정도의 구실은 할 수 있다. 그렇다면 우리가 규칙과 기호를 쉽게 포기하지 못하는 것은 실용적인 까닭에서이다. 뉴우턴 역학과 양자 역학의 비유를 가지고 설명해 보자면 우선은 양자적인 접근을 직접적으로 하기에는 곤란한 많은 고전적인 현상에 대해 뉴우턴 역학이 설명-더 정확히 말하자면 대략적인 설명-을 할 수 있기 때문이다. 그리고 또 역사적으로 볼 때 뉴우턴 역학은 양자 역학의 원리들을 발견하는 데 꼭 필요한 길잡이 구실을 했다는 것을 인정하기 때문이다. 상위차원에 있는 이론의 원리들을 길잡이로 삼지 않고서는 하위 차원의 이론은 어두운 길을 헤매고 있을 수 밖에 없을 것이다. 이런 필요성은 현재 연결주의가 처한 상황을 생각해 본다면 더 절실해진다. 스몰렌스키에 따르면 연결주의는 고전주의로 말하자

면 아리스토텔레스 정도의 수준밖에 와있지 못하다(Smolensky, 1991, 202쪽).

거듭 말하지만 이제 막 피어나는 연결주의에도 의미있는 인지 이론의 기회를 주어야 한다. 인지에 관한 이론을 경험에 토대를 두고 쌓아가려는 그 이론에게 단지 선행적인 이유만으로 인지 이론의 길을 막는 것은 아무리 봐도 좋아 보이지 않는다.

한 가지 추측이긴 하지만 만약 인지과정에서 기호와 규칙의 필요성이 단순히 실용적인 근거에 불과하다면 고전주의와 연결론은 처치랜드(1986)가 인간 심리학과 신경 생리학에 대해서 말한 '함께 진화하기'(co-evolution)의 운명을 걸을지도 모른다는 생각이 든다. 그래서 현재에는 고전주의가 연결주의와 갖는 잠정적인 관계를 통해 그 길잡이로서 갖는 필요성이 요구되지만 어느 정도 이론의 성숙이 이루어진 단계에서는 완전한 환원과 제거가 이루어질지도 모른다. 그러나 정말 그런 제거가 이루어질지 아닐지는 어디까지나 경험적인 문제라는 걸 지적하며 이 글을 맺어야겠다.

참 고 문 헌

- 최 훈(1993), "라이벌이나 구현이나?", HCI 연구실·기초이론부(편역), 『HCI 기초 이론』, 2호, 서울: 29-44쪽.
- 최 훈(1994a), 「연결주의와 구성적 구조: 포더와 그의 동조자들에게 보내는 응답」, 서울대학교 철학과 석사학위 논문.
- 최 훈(1994b), "한갓 구현이 아닌 구성적 구조", HCI 연구실·기초이론부(편역), 『HCI 기초 이론』, 3호, 서울.
- Bechtel, W.(1988), "Connectionism and Philosophy of Mind: An Overview", in W. Lycan(ed.)(1990), *Mind and Cognition: A Reader*, Oxford: Basil Blackwell: 252-273.
- Bechtel, W. and A.A. Abrahamsen, (1991), *Connectionism and the Mind: An Introduction to Parallel Processing in Networks*, Basil Blackwell.
- Butler, K.(1993a), "Connectionism, Classical Cognitivism and the Relation between Cognitive and Implementational Levels of Analysis", *Philosophical Psychology* 6: 321-333.

- Butler, K.(1993b), "On Clark on Systematicity and Connectionism", *British Journal for the Philosophy of Science* 44: 37-44.
- Chalmers, D.J.(1990), "Syntactic Transformations on Distributed Representations", *Connection Science* 2: 53-62.
- Chalmers, D.J.(1993), "Connectionism and Compositonality: Why Fodor and Pylyshyn Were Wrong", *Philosophical Psychology* 6: 305-319.
- Churchland, P.M.(1988), *Matter and Consciousness*, (revised ed.), Cambridge, MA: MIT Press/Bradford Books.
- Churchland, P.M.(1989), *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge, MA: MIT Press/Bradford Books.
- Churchland, P.S.(1986), *Neurophilosophy: Toward a Unified Understanding of the Mind-Brain*, Cambridge, MA: MIT Press/Bradford Books.
- Churchland, P.S. and T.J. Sejnowski, (1989), "Neural Representation and Neural Computation", in W. Lycan(ed.)(1990), *Mind and Cognition: A Reader*, Basil Blackwell: 224-252.
- Clark, A.(1989), *Microcognition: Philosophy, Cognitive Science and Parallel Distributed Processing*, Cambridge, MA: MIT Press/Bradford Books.
- Clark, A.(1991), "Systematicity, Structured Representations and Cognitive Architecture: A Reply to Fodor and Pylyshyn", in T. Horgan and J. Tiensen(eds.)(1991), *Connectionism and the Philosophy of Mind*, Dordrecht: Kluwer Academy Publishers: 198-218.
- Cummins, R.(1983), *The Nature of Psychological Explanation*, Cambridge, MA: MIT Press/Bradford Books.
- Davidson, D.(1965), "Theories of Meaning and Learnable Languages", in his(1984) *Inquiries into Meaning and Truth*, Oxford: Oxford Univ. Press: 3-15.
- Dennett, D.C.(1986), "The Logical Geography of Computational Approaches: A View from the East Pole", in M. Harnish and M. Brand(eds.), *Problems in the Representation of Knowledge*, Univ. of Arizona Press: 59-79.
- Dennett, D.C.(1991), "Mother Nature Versus the Walking Encyclopedia: A Western Drama", in W. Ramsey, S.P. Stich, and D.E. Rumelhart(eds), *Philosophy and*

- Connectionist Theory*, NJ: Erlbaum: 21-30.
- Flanagan, O.(1991), *The Science of the Mind*, (2nd ed.), Cambridge, MA: MIT Press/Bradford Books.
- Fodor, J.(1975), *The Language of Thought*, New York: Thomas Y. Crowell.
- Fodor, J.(1987), *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: MIT Press/Bradford Books.
- Fodor, J.(1985), "Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum", *Mind* 94: 76-100.
- Fodor, J.A and McLaughlin, B.P.(1991), "Connectionism and the Problem of Systemicity: Why Smolensky's Solution Doesn't Work", in T. Horgan and J. Tiensen(eds.)(1991), *Connectionism and the Philosophy of Mind*, Dordrecht: Kluwer Academy Publishers: 331-354.
- Fodor, J.A. and Pylyshyn, Z.W.(1988), "Connectionism and Cognitive Architecture: A Critical Analysis", *Cognition* 28: 3-71.
- Goschke, T. and Koppelberg, D.(1990), "Connectionist Representation, Semantic Compositionality, and the Instability of Concept Structure", *Psychological Research* 52: 253-270.
- Haugeland, J.(1985), *Artificial Intelligence: The Very Idea*, Cambridge, MA: MIT Press/Bradford Books.
- Hawthorne, J.(1989), "On the Compatibility of Connectionist and Classical Models", *Philosophical Psychology* 2: 5-15.
- Hinton, G.E., McClelland, J.L. and Rumelhart, D.E.(1986), "Distributed Representations", in D.E. Rumelhart and J.L. McClelland and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructures of Cognition, vol. 1: Foundations*, Cambridge, MA: MIT Press/Bradford Books: 77-109.
- Horgan, T. and J. Tiensen(eds.)(1991), *Connectionism and the Philosophy of Mind*, Dordrecht: Kluwer Academy Publishers
- Kaye, L.J.(1993), "Semantic Compatibility: Still the Only Game in Town", *Analysis* 53: 17-23.
- Lahav, R. (1989), "Against Compositionality: the Case of Adjectives", *Philosophical Studies* 57: 261-279.
- McClelland, J.L. and Rumelhart, D.E.(1981), "An Interactive

- Activation Model of Context Effects in Letter Perception, Part I: An Account of the Basic Findings”, *Psychological Review* 88: 375-407.
- McLaughlin, B.(1993), “The Connectionism/Classicism Battle to Win Souls”, *Philosophical Studies* 71: 163-190.
- Newell, A.(1980), “Physical Symbol Systems”, *Cognitive Science* 4: 135-183.
- Pylyshyn, Z.W.(1984), *Computation and Cognition: Toward a Foundation for Cognitive Science*, Cambridge, MA: MIT Press/Bradford Books.
- Pylyshyn, Z.W.(1989), “Computing in Cognitive Science”, in M. I. Posner(ed.), *Foundations of Cognitive Science*, Cambridge, MA: MIT Press/Bradford Books: 51-91.
- Ramsey, W., S. Stich, and J. Garon(1990), “Connectionism, Eliminativism, and the Future of Folk Psychology”, in D. J. Cole, J. H. Fetzer, and T. L. Rankin(eds)(1990), *Philosophy, Mind, and Cognitive Inquiry: Resources for Understanding Mental Processes*, Dordrecht: Kluwer Academy Publishers: 117-144.
- Ramsey, W., S.P. Stich, and D.E. Rumelhart(eds.)(1991), *Philosophy and Connectionist Theory*, NJ: Erlbaum.
- Rumelhart, D.E. and McClelland, J.L. (1986), “PDP Models and General Issues in Cognitive Science”, in D.E. Rumelhart and J.L. McClelland and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, vol. 1: Foundations, Cambridge, MA: MIT Press/Bradford Books: 109-146.
- Schiffer, S.(1987), *Remnants of Meaning*, Cambridge, MA: MIT Press/Bradford Books.
- Schiffer, S.(1991), “Does Mentalese Have a Compositional Semantics?”, in B. Loewer and G. Rey(eds,)(1991), *Meaning in Mind: Fodor and his Critics*, Oxford: Basil Blackwell: 181-199.
- Schiffer, S.(1993), “Compositional Supervenience Theories and Compositional Meaning Theories”, *Analysis* 53: 24-29.
- Smolensky, P.(1987), “The Constituent Structure of Connectionist Mental States: A Reply to Fodor and Pylyshyn” in T. Horgan and J. Tiensen(eds.)(1991), *Connectionism and the*

- Philosophy of Mind*, Dordrecht: Kluwer Academy Publishers: 281-308.
- Smolensky, P.(1988), "On the Proper Treatment of Connectionism", in D. J. Cole, J. H. Fetzer, and T. L. Rankin(eds)(1990), *Philosophy, Mind, and Cognitive Inquiry: Resources for Understanding Mental Processes*, Dordrecht: Kluwer Academy Publishers: 145-206.
- Smolensky, P.(1990), "Tensor Product Variable Binding and the Representation of Symbolic Structures in Connectionist Systems", *Artificial Intelligence* 46: 159-216.
- Smolensky, P.(1991), "Connectionism, Constituency, and the Language of Thought", in B. Loewer and G. Rey(eds.) (1991), *Meaning in Mind: Fodor and his Critics*, Oxford: Basil Blackwell: 201-228.
- Tye, M.(1991), "Representation in Pictorialism and Connectionism", in T. Horgan and J. Tiensen(eds.)(1991), *Connectionism and the Philosophy of Mind*, Dordrecht: Kluwer Academy Publishers: 309-330.
- van Gelder, T.(1989), *Distributed Representation*, Doctorial Dissertation, University of Pittsburgh.
- van Gelder, T.(1990), "Compositionality: A Connectionist Variation on a Classical Theme", *Cognitive Science* 14, 353-384.
- van Gelder, T.(1991a), "Classical Questions, Radical Answers: Connectionism and the Structure of Mental Representations", in T. Horgan and J. Tiensen(eds.)(1991), *Connectionism and the Philosophy of Mind*, Dordrecht: Kluwer Academy Publishers: 355-379.
- van Gelder, T.(1991b), "What is 'D' in PDP? A Survey of the Concept of Distribution" in W. Ramsey, S.P. Stich, and D.E. Rumelhart(eds), *Philosophy and Connectionist Theory*, NJ: Erlbaum: 33-59.